

73283

STIC-Biotech/ChemLib

From: Rao, Manjunath N.  
Sent: Wednesday, August 14, 2002 12:47 PM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 10/003,759

From: Manjunath N. Rao  
Art Unit 1652, Room 10A11  
Mail Box in Room 10C 01  
Phone: 306-5681

Date: 8-14-02

Please search the following as soon as possible for application with serial number **10/003,759**

**SEQ ID NO: 1, SEQ ID NO:3 and nucleotides 52-783 of SEQ ID NO:3** against all commercial nucleic acid databases including issued patents database and pending application database and provide a **print of all results**.

**SEQ ID NO: 2 and amino acids 40-261 of SEQ ID NO:2** against all commercial protein databases including issued patents database and pending application database and provide a **print of all results**.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.  
Biotechnology Patent Examiner  
Art Unit 1652, Room 10A11  
Crystal Mall 1, USPTO.

Searcher: Manjunath Rao  
Phone: 306-4501/4506  
Location: Biotech Lab  
Date Picked Up: 8/15/02  
Date Completed: 8/20/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: 10 min  
Online time: 9 min

TYPE OF SEARCH:  
NA Sequences: 3  
AA Sequences: 2  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: ABSS02/03  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 09:50:56 ; Search time 6219.51 Seconds

(without alignments)  
6315.467 Million cell updates/sec

Title: US-10-003-759-1

Sequence: 1 999cgagtcgaacgtcgggc.....tctccgacactacaacac 1877

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hlg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pr: \*  
9: gb\_ro: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_hum: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hlg\_hum: \*  
31: em\_hlg\_inv: \*  
32: em\_hlg\_other: \*  
33: em\_hlgo\_inv: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Match Length	ID	Description
------------	-------	--------------	----	-------------

1	1877	100.0	1877	6	AX339680	AX339680 Sequence
2	1837.8	97.9	1874	1	RM072637	U72637 Rhodothermus
3	786	41.9	786	6	AX339682	AX339682 Sequence
4	131.8	7.0	1116	1	AF233376	AF233376 Streptomy
5	131.8	7.0	1116	6	ARI29926	ARI29926 Sequence
6	131.8	7.0	1116	6	ARI68360	ARI68360 Sequence
7	105.8	5.6	1716	6	ARI68362	ARI68362 Sequence
8	105.2	5.6	1944	1	AF145250	AF145250 Rhodother
9	97.8	5.2	1357	1	SHU51222	U51222 Streptomyce
10	96.8	5.2	1470	1	SREGLS	X73953 S.rochei eg
11	89	4.7	1311	1	AF130408	AF130408 Streptomy
12	87.6	4.7	1565	1	SL004629	U04629 Streptomyce
13	87.6	4.7	11782	1	SCG11A	AL133210 Streptomy
14	69	3.7	11202	1	AE005123	AE005123 Halobacte
15	68.8	3.7	1262	1	AF355723	AF355723 Burkholde
16	68	3.6	125020	9	AF429315	AF429315 Homo sapi
17	66	3.5	14483	1	AE006992	AE006992 Mycobacte
18	66	3.5	67200	1	MTV017	AL021897 Mycobacte
19	63.2	3.4	125020	9	AF429315	AF429315 Homo sapi
20	57.6	3.1	1983	6	AX047403	AX047403 Sequence
21	56.6	3.0	45313	1	SCD95A	AL57432 Streptomy
22	56.6	3.0	201050	1	AL646064	AL646064 Ralstonia
23	55	2.9	36368	1	SC9B5	AL035206 Streptomy
24	54.8	2.9	346887	1	AP002995	AP002995 Mesorhizo
25	54.4	2.9	38995	1	SCF34	AL109974 Streptomy
26	54	2.9	34182	1	SC111	AL096849 Streptomy
27	53.6	2.9	77534	1	AF235504	AF235504 Streptomy
28	53	2.8	33779	1	SCGD3	AL096882 Streptomy
29	52.2	2.8	1849	1	SGSTRB1F	X78973 S.galbus DS
30	52	2.8	6530	1	AB017438	AB017438 Streptomy
31	52	2.8	26500	1	SC9H11	AL356592 Streptomy
32	52	2.8	349116	1	AP003003	AP003003 Mesorhizo
33	51.4	2.7	3644	5	AF034576	AF034576 Gallus ga
34	51.4	2.7	10029	1	AF263012	AF263012 Streptomy
35	51.2	2.7	22115	1	SCC22	AL096838 Streptomy
36	51	2.7	77457	1	AF210249	AF210249 Streptomy
37	50.6	2.7	5525	1	SGSTRB1F	X78974 S.glaucisce
38	50.6	2.7	14596	1	AE007168	AE007168 Mycobacte
39	50.6	2.7	25459	1	SGA6985	AJ006985 Streptomy
40	50.6	2.7	30302	1	MTGCG11	292774 Mycobacteri
41	50.6	2.7	40806	1	MSGY23	AD000016 Mycobacte
42	50	2.7	14939	1	AE005092	AE005092 Halobacte
43	50	2.7	29625	1	SCR22A	AL159178 Streptomy
44	50	2.7	36394	1	SCD35	AL160312 Streptomy
45	50	2.7	40104	1	SCF11	AL132662 Streptomy

## ALIGNMENTS

RESULT 1	AX339680	1877 bp	DNA	linear	PAT 10-JAN-2002
LOCUS	AX339680	Sequence 1 from Patent WO0196382.			
DEFINITION	AX339680				
ACCESSION	AX339680				
VERSION	AX339680.1	GI:18135682			

## KEYWORDS

## SOURCE

## ORGANISM

Rhodothermus marinus.  
Rhodothermus marinus

Bacteria; CFB group; Rhodothermus group; Rhodothermus.

1 (sites)

Wicher, K.B., Holst, O.P., Hachem, M.Y., Karlsson, E.M. and

Hreggvidsson, G.O.

Thermotable cellulase

Journal: J Biol Chem 273:12055-12062 (1998)

Prokaryotic cell

## CDS

Location/Qualifiers  
1..1877  
/organism="Rhodothermus marinus"  
/db\_xref="taxon:29549"  
710..1495  
/note="unnamed protein product"  
/codon\_start=1

```
/transl_table=11
/protein_id="CAD20684.1"
/db_xref="GI:18135683"
/translation="MNMRAVLVLSLLFLFGCDMLFPDDDNKKEPEREPRVELCGR
WQARDVAGRYRIVNNWGAETACILEVELETGNFTITPADNDNGNNVAIPAIFGC
HMGACTSNGSLPRRVOELSDVRSMTLPITTRGRNMAADIDMSPTNNVANSYGAE
LMIMLMNGGVMDGSGSRVATVELAGATMEVNYADMOMNI IAVRITPTPTVSSELDLKA
FIDDAVARGYRREPMTHAVETFEIWEAGLRSDFSVTQ"
```

BASE COUNT 331 a 582 c 632 g 332 t

Query Match 100.0%; Score 1877; DB 6; Length 1877;  
Best Local Similarity 100.0%; Pred. No. 3.6e-278;  
Matches 1877; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gggcagtcgaacgctcgccgagcttcgacagtcgaacgagcttccacgctgaltgcg 60  
DB 1 GGGCAGTCCGACGCTCGCGCGCTCGCAGATCGAAGCGCTCTCCACGCTGATTCCGC 60  
QY 61 tggctggcagcagcagctggcgctggcgagcttcccttcaacttctgcagattggccct 120  
DB 61 TGGCTGGCAGCAGCTGGCGCTGGCGAGATTTCCTTCTACTTCCGTCAGATTGGCCCT 120  
QY 121 tacgactacgagcagcagccgcaacgagcgtgtcccggaagcagcagctgcacacatg 180  
DB 121 TACGACTACGGCGCAGAGCCCGCACAGCGTGTCCGGGAAGCGCAGTTGGCCACCATG 180  
QY 181 tggcgcgctctcaatgaggtatgtgttgacacacagatgtggcgagacacacaccta 240  
DB 181 TGGCGGTCTCTACATACGGGATGCTGTGACACCGATGTGGGGAGACACACACTA 240  
QY 241 ccttcgagcagcaacgctgaagctggcagcagcgttcgttcggcgcttcgcaacagcta 300  
DB 241 CCTTCGGCGGAGACACTGAAGTGGCGAGCGCTCGGCGCTTGGCGCTGGCCAAAGCTA 300  
QY 301 cggcttcgagatgtgttctactccggtlcccatctacgaacgcatggaacgtgagatgg 360  
DB 301 CGGCTTCGGGAGTGTCTACTCCGCTCCCATCTACGAACCGCATGTGAGAGTGG 360  
QY 361 ccgcttcgagctctattctcgctatgacgaagggagactgtctctggcgcccgcgagcgg 420  
DB 361 CCGCTTCGGCTCTATTTCGCTATGCCGAAGGGGAGCTGTCTGGCGCCCGCCGACGG 420  
QY 421 cggcgagagctcgtgtatctggagaccggaacgctgtcttccacccggcgcggttacggt 480  
DB 421 CGGCGGAGAGTGTGATTCGGGGACCGGACCGGTCTTCACCCGGCGGGGTACGGGT 480  
QY 481 cgaagaggaacgctgtgtgtgtgtgagatcccgcgcttccgactccgagcggtggccta 540  
DB 481 CGAAGGCGAAACGCTGTGTGTGTGTGAGTCCGCGCGTCTCGATCCGAGCGGTGCGCTA 540  
QY 541 tggctgtgtccaatacgcgcagcagcagcgtgttcaacccggcgcgcttcgcgcttcgac 600  
DB 541 TGGCTGTGTCCAATACGCGCGACGCGCATGTTCAACCGGGCGCGCTTGGCCGCTTGGCC 600  
QY 601 ctctcgcagcagcagcagctggcggaagcagctgagcagcagcaacgggtgtcttgcatgca 660  
DB 601 CTTCCGACCGCAGCACTGGCGCGGAAGCGACTGAGCGACGCAACCGGTGTCTTGCATCGCA 660  
QY 661 caaggagactctcfacacttgaatgtgacctgtgtcatctcaatgaatgaatgaacgcat 720  
DB 661 CAGGGGCACTTCGTACTTGAAAGTGCCTTTGTATTTCAATGGAATGAAGAGTCAAT 720  
QY 721 gcgctggcagtaactgtccctggacgttcatctgtctgttgatggcagctgttccgga 780  
DB 721 GCGTGGCGTAAGTGTCTGAGCGCTGTATTGTGCTTTTGGATGGCAGCTGGCTTTCGGA 780  
QY 781 tggcagacacgaaggaaggaacggagccttgagcccgagccgacgctgagctgttgagag 840  
DB 781 TGGCAGACACGGAAGGAACCGGAGCCTGAGCCCGACGACCGTTCAGAGCTGTGCGGAGC 840  
QY 841 ctggagagcgcgcgctgttgccggggggcgctaacgggtgatacaacagctatggggcg 900

```
|||||  
DB 841 CTGGAGCGCGCGAGTGTGGCCGGGGCGGTACCGGGTGAATCAACAGTATGGGGCCG 900  
QY 901 ggaagacggcccgatgtatgtgaggttgagacttgaaagggcaacttccatcaacacggcg 960  
DB 901 GGAAGACGGCCCGATGTGATGTGAGGTGCGACTGGAAAGGGCACTTCACATCAACAGGGCG 960  
QY 961 cga tcacgaacgaagcaacgaacgtggcgccctatccggccatctacttccgggtggccactg 1020  
DB 961 CGATCAAGCAACAGGCAACACTGGCCGCTATCCGGGCATCTACTTGGGGTGGCAGCTG 1020  
QY 1021 gggcgcttcagcagcaaatccggatctgcggcgcgctgtcagagagctgtlccgaagctgcg 1080  
DB 1021 GGGCGCTTCAGCAGCAATTCGGGATTCGGCGGCGCTGTGAGAGCTGTCCGACGCTGGC 1080  
QY 1081 caagacttgagcgttaacgcgcga tcacgagcgggcgcttgaaatgcgcgcttaagaaactgt 1140  
DB 1081 CAGAGCTTGAGCGCTACCGCCGATCAAGAGGCGCCTGTGAATGCCGCTTACGACACTGTG 1140  
QY 1141 gtccag tcccgtaacgaatlcgcggcaacggtctacagcggcgcgccgagctgaltgaltgt 1200  
DB 1141 GTTCAGTCCCGTACGAATTCGGGCAACGGGTACAGCGCGCGCGGCTGATGATCTG 1200  
QY 1201 gctgaacttggaacggcgcgctgtatgcggcgcgagcggctgtggccacggctggaactgtgc 1260  
DB 1201 GCTGAACCTGGAAAGCGCGGCTGTATGCCGGCGGCGCACCCGCTGTGGCAACTGGC 1260  
QY 1261 cggcgccaccttggaagcttgatgtccgacttgagacttgaaatctacatccgcttaccggcg 1320  
DB 1261 CGGGCGCACTGGGAAGTCTGATGCTGATCCGACTGTGAGACTGTGAATTCATGCGCTACCGGGG 1320  
QY 1321 cagcagcccaacccacgctcggltgagcagcagctgagccttgaaaggccttca tcgaacgaagcgt 1380  
DB 1321 CACGAGCGCCACCAACGCTCGGTGAGCAGCTGAGCTGAAGAGCGCTTCATGACAGACCGGT 1380  
QY 1381 cggcccggtctataatccggtccgaggtgtatctgca tgggtggagagcggtcttgaact 1440  
DB 1381 CGCCCGCGCTTACATCCGCGGAGTGTGATCTGATGCGGTGAGAGCGGCTTCCAACTT 1440  
QY 1441 ctggagggcgggcgcgctgtcgagcagccga tttctcgtlaacggtgcaatgaagctgt 1500  
DB 1441 CTGGAGGGGGGGGCGCGGTGTGCGAAGCGCGCATTTTCGTATAGCGTACAGTACGCTGT 1500  
QY 1501 cacaagggcaacagcgttaagcagaagaacacggctcggggcgctatgtcggggcggt 1560  
DB 1501 CACACGGGCGCACAGCGCTAGGCCAGAGAGCACCGCTCGGGCGCTTATCGGGCGCGGT 1560  
QY 1561 cgtcttgctcgaacttggtagtgcgtacgagggcgctcagccggcgctgtgtgttccg 1620  
DB 1561 CGCTTTGTGCTTACCTTTGTAGTGCCTTACGAGGGGCTACGGCGGGTGTGGTTCCTCC 1620  
QY 1621 tggagagcgcttcgttcgtgcggacgaagacttaccgaaaagcaaggagagacgtg 1680  
DB 1621 TGGAGAGCGCTTCTGTTCTGTCGCGAGAGAACTTATCCGAAAGCAAGGAGAGTCTGG 1680  
QY 1681 taagcaagggcttgagcttgagcagcagaagagcgaggttggtagtaagcagaagaggt 1740  
DB 1681 TAAGCAAGGGCTTTGAGCTGCGCACAGAGCAAGGGGTTTGTATGACCAAGCAAGCGGT 1740  
QY 1741 ccgcggaagtggcctcaaccccgagcagcacttcgctgcgtcccgcttccgtactacagag 1800  
DB 1741 CCGCGGAAGTGGGCTCAACCGCGAGCGCACTTGTGCTGGCGCTCGTCTGACTACAGAGG 1800  
QY 1801 cgttggcgtgtcttcaaaagcgagcgtataaagctataaagcttcaagctcgaacttcc 1860  
DB 1801 CGTTGGCGCGTGTCAAAACCGCAGGCTATTAAGCTATACAGGTGGAATTCGCAAACTTCT 1860  
QY 1861 ccgacacttaacaaacac 1877  
DB 1861 CCGACACTTACAAACAC 1877
```

RESULT 2



RMU72637	1874 bp	DNA	linear	BCT 11-MAY-1999
DEFINITION	Rhodothermus marinus cellulase (celA) gene, complete cds.			
ACCESSION	U72637			
VERSION	U72637.1			
KEYWORDS	GI:2304960			
SOURCE	Rhodothermus marinus.			
ORGANISM	Rhodothermus marinus.			
REFERENCE	Bacteria: CFB group: Rhodothermus group: Rhodothermus.			
AUTHORS	1 (bases 1 to 1874)			
TITLE	Hallidorsdottir, S., Thorolfsson, E.T., Spilliaert, R., Johannsson, M., Thobbjarnardottir, S.H., Palsdottir, A., Hregeyldsson, G.O., Kristjansson, J.K., Holst, O., and Eggertsson, G. Cloning, sequencing and overexpression of a Rhodothermus marinus gene encoding a thermostable cellulase of glycosyl hydrolase family 12			
JOURNAL	Appl. Microbiol. Biotechnol. 49 (3), 277-284 (1998)			
MEDLINE	9824232			
REFERENCE	2 (bases 1 to 1874)			
AUTHORS	Hallidorsdottir, S.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-SEP-1996) Department of Molecular Biology, Institution of Biology, Grensasvegur 12, Reykjavik 108, Iceland			
FEATURES	Location/Qualifiers			
source	1..1874			
gene	/organism="Rhodothermus marinus"			
CDS	/strain="T1378"			
	/db_xref="taxon:29549"			
	710..1493			
	/gene="celA"			
	710..1492			
	/gene="celA"			
	/BC_number="3.2.1.4"			
	/note="belongs to family H of cellulase catalytic domains; endo-1,4-beta-glucanase"			
	/codon_start=1			
	/transl_table=11			
	/product="cellulase"			
	/protein_id="AAB65594.1"			
	/db_xref="GI:2304961"			
	/translation="MNMRALVLSLLFLFGCDMLFPDGDNGKEPEPEPEVELGGR HDARVAGRRYRVINNVGAFEAOCIEVELETFENFTIRADHDNGNNVAAYPAIFGC HMAPARTRDCAARAGAVRAHELDVPTPTGRMNAAYIWMSPYNSNGISGAEF MTLWNNGSVMPGSGSRVATVELAGATWEVYADMDNNTATRTPTTSVSELDLKAFL IDDAVARGIRREWYLHAETGFELEGGAGLRADFSVTVO"			
BASE COUNT	331 a	580 c	631 g	332 t
ORIGIN				
Query Match	97.9%	Score 1837.8	DB 1	Length 1874
Best Local Similarity	99.7%	Pred. No. 3.6e-272		
Matches 1872; Conservative	0	Mismatches 2	Indels 3	Gaps 3

QY	301	cggctccggaatggtgctactccggtcccatctacgaacgcatggaactggaatg	360
DB	301	CGGCTCCGGAGTGTGATCTACCTCCGTCACATCTACGAACGCATGGAACGTGAGATGG	360
QY	361	ccgctcggtgtctattccgcatgacgaagggagactggtctcgcccgccgaag	420
DB	361	CCGCTCGGCTCTATTTCGCTATGCCGAAGGGGACGTGCTCCGCCGCCGACGG	420
QY	421	cggcggaattcggtatgacggagcgaacggtgtctccaccgagcggtgaaggt	480
DB	421	CGCGCGGAGATTGATATTCGGGACCGGACCGCTGCTTCCACCCGGCGGGTACGGGT	480
QY	481	cgaagcggaacgctggtggtgagtcgacgagtcgcggtctccgacgcgagcggtg	540
DB	481	CGAAGCGGAACGCTGATGATGAGTCGCGCGCTCTCCGATCCGACGGGATGGCTA	540
QY	541	tggctgtccaaatgacccgacgacgctgttcaaccgggagccgctccgctccg	600
DB	541	TGGCTGTCCAAATGACCCGACGACGCTGTTCACACCGGCGCTCGCCCTTCGCGC	600
QY	601	cttcgacacgacgactggcggaagcgactgagcgacgcaacggtctgcatgga	660
DB	601	CTTCGACACGACGACTGGCGGAAGCGACTGAGCGACGCAACCGGCTTGCATGCCA	660
QY	661	caggggcacttcgtacttgaagtgccttgcatttcaatggaataatgaactcat	720
DB	661	CAGGGCACTTCTACTTGAAGTGCCTTGTTCATTTCAATGAATGAATGAACGTCA	720
QY	721	ggctggtgatactgttcctgagcctgttatgtcgtttgtagtgactggtcttccga	780
DB	721	GCGTGGTGAATGCTGCTCTGACCTGTATTGCTGTTGATGTCGACTGCTCTTCCCA	780
QY	781	tggcgcaacggaagaaacggaagcctgagcccgagccgagcctgagcgtgagcg	840
DB	781	TGGCGCAACGGAAGAAACCGGAGCCTAGCCCGAGCCGACCGCTGAGCTGCGGAGG	840
QY	841	ctggagacgagcgatgctgagcgaggcgactacgagtgatacaacaagctatggg	900
DB	841	CTGGAGACGAGCGATGCTGAGCGAGGCGCTACCGGATGATCAACAGCTATGGGGCGC	900
QY	901	ggagaccccgatgcatgagtgagtcgagctggaagacgggaacttacaacagagcg	960
DB	901	GGAACCCCGCATGATGAGTGAAGTCCGACTGGAAACGGGCACTTACATCACACGGGC	960
QY	961	cgaatcaagaacggaacacgctggccgactaaccggacatctactcgggtgacact	1020
DB	961	CGATCAACGAACGGAACACGCTGGCGGCTATCCGGCATCTACTTGGGTCACCT	1019
QY	1021	ggcgccctgcaagacgaatctggatgtccgagcgagcggtgacgagctgcaagctg	1080
DB	1020	GGCGCGCTGCACGAGCAATTCGGGATTTG-CGCGGCCGTCACGAGCTGTCCACGTTGGC	1078
QY	1081	caagagctgagcgtacagcgcgatcagcagagcgagctggaatgacgacctagaactctg	1140
DB	1079	CACGAGCTGAGCG-TCACGCCGATTCACGACGGCCGCTGGAATGCCGCTACGACATCTG	1137
QY	1141	gttcaatcccgatcaagaattccggcaacggttacaagcgagcgagcgagctgatactg	1200
DB	1138	GTTCAATCCCGATCAGCAATTCGGCAACGGCTACAGCGGGCGCGAGCTGATGATCTG	1197
QY	1201	gctgaactgaaacgagcggtgatactgcggcgagcgacgacggttgccacgctggaactg	1260
DB	1198	GCTGAATGGAACGAGCGGCTGATGCTCGGGCGGACCCGCTGAGCCACCTGGAATGTCG	1257
QY	1261	cgggacacactgggaagctggtatgacgacgtggagactggaattacatcgctaccggcg	1320
DB	1258	CGGGACACTGGGAAGCTGTGATATGCCACTGGGACTGGAATTAATCGCTTACCGGGG	1317
QY	1321	caagacgcccacacgctcggtgagcgagctggaactggaagcgcttaccgaacgagcg	1380
DB	1318	CACGACGCCACACGCTCGGTGAGCGAGCTGGAAGCGCTTCATCGACGACCGGCT	1377
QY	1381	cgcccgcgctataccgagcgaggtggtatctgcatgctggtggtggaacgggtctggaact	1440

```
|||||
Db 1378 CGCCCCGCGCTACATCCGCGGAGTGTATGTCATCGCGGTGAGACGGCGCTTCCAAAGT 1437
Qy 1441 cggggagggcggggcccggctctcgagagcgccgaattttccgtgaagcgctgaagcctgt 1500
Db 1438 CTGGGAGGGCGGGGCGGCTCTGGGAAGGGCCGATTTTTCGTAAACGCTGCACAGTACGCTGT 1497
Qy 1501 caacagggagcaacagctgaagccagagaagcaaccgctcggggcgctctatgcggggcgagc 1560
Db 1498 CACACGGGACACGAGCTTAGGCCAGAGAACGCCCTCGGGGCGCTTATGCGGGCGGCGCT 1557
Qy 1561 cgccttgctgaccttgatgctgctacgagagcgctcgaagccggcgctgctgcttccc 1620
Db 1558 CGCTTTGTCTGACTTGTAGTGTGAGTGCCTACGAGGCGTACGCGCGCGGTGGTGGCTTCCC 1617
Qy 1621 tggagggcgcttgcttcgctgagcgagagactatccgaaaaagagagagagactctg 1680
Db 1618 TGGAGGCGCTTCTGTTCTGTCGCCGACGAGAACTTATCCGAAAGGAGAGACCTTG 1677
Qy 1681 taagccagggccttgctgactgagcagcagagcaggggttctgtagtaagccagagagcgt 1740
Db 1678 TAAGCCAGGAGCTTTGACATGCGGACAGCAAGCAGGGCTTGTATGATGACGACGAGAGCGT 1737
Qy 1741 cggcgagagtgagcctcctaccccgagagcacttcgctgctgctcgtgctgactacgagcg 1800
Db 1738 CGCGCAAGTGGCTTCCACCGGACGCGACCTTCGCTGCGCTCGCTGACTACGACGAGC 1797
Qy 1801 cgttgagcgctgctcgaacgagcagctataaagctatagagctgagatcgagagcctct 1860
Db 1798 CGTTGGCGCTGCTTCAACGCGGCGGCTATTAAGCTATACAGCTGCAAAATCGCAACTTCT 1857
Qy 1861 ccgacacttaacaaac 1877
Db 1858 CCGACACTTACAAACAC 1874

RESULT 3
LOCUS AX339682 786 bp DNA linear PAT 10-JAN-2002
DEFINITION Sequence 3 from Patent WO0196382.
ACCESSION AX339682
VERSION AX339682.1 GI:18135684
KEYWORDS
SOURCE Rhodothermus marinus.
ORGANISM Rhodothermus marinus
REFERENCE 1 (sites)
AUTHORS Wicher, K.B., Holst, O.P., Hachem, M.Y., Karlsson, E.M. and Hreggvidsson, G.O.
TITLE Thermostable cellulase
JOURNAL Patent: WO 0196382-A 3 20-DEC-2001;
FEATURES
    source 1. 786
    /organism="Rhodothermus marinus"
    /db_xref="taxon:29549"
BASE COUNT 141 a 236 c 279 g 130 t
ORIGIN
```

Query Match 41.9%: Score 786; DB 6; Length 786;  
Best Local Similarity 100.0%; Fred. No. 4.4e-111;  
Matches 786; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 710 atgaacgtatgctgctgactgctgctgagcctgctatgctgctgctgagtcgagtcg 769
Db 1 ATGAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
Qy 770 cttcttccgattgagcaacaggaagaaacggagcctgagcccgagccgagccgagtcgag 829
Db 61 CTTCTTCCGATTGAGCAACAGGAAGAAACCGAGCCTGAGCCGAGCCGAGCCGTCGAG 120
Qy 830 cgtgcgagagcgtggaagcgcgagatgagccgggggagcgtacccgggtgatcaacaaac 869
```

```
|||||
Db 121 CTGTGCGGAGCGCTGGGAGCGCGCGCATGTGCGGGGGGCGCTACCGGTGATCAACAAAC 180
Qy 890 grattggggcgaggagacgcgcacagtgatgagtgctgagcttgagaaagcgcaacttaag 949
Db 181 GATATGGGGGCGGAGAGCCCGCCAGTGCATTTAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 240
Qy 950 atcacagggcgagatcacagcaacagcgaacagctgagccctatccgcaacttacttc 1009
Db 241 ATCACAGGGCGGATCACAGCAACAGCAACAGTGTGCGGCGCTATTCGCGCATCTTACTTC 300
Qy 1010 ggggtgcaactggggcgctgcaagagcaatctgggattgctgagcgcgctgcaagagctg 1069
Db 301 GGGTGCACATGGGGCGCTGTCACAGCAATTCGGGATTCGCCGCGCGCTGACGAGGCTG 360
Qy 1070 tccgagtgctgcaagcgtgagcgtcgaagccgagatcacagagggcgccgtggaatcgccg 1129
Db 361 TCCGAGTGTGCGGACGAGCTGAGAGCTCACGCGCATACAGACGCGCGCTGGAATGCGCGC 420
Qy 1130 taagacatctggttcaagctccgctcaagaaatcccgcaacgagctacagcgcgccgag 1189
Db 421 TAGGACATCTGTTGATGATCGCTCCGTCACGAATTCGCGCAACGGGTACAGCGCGCGCGAG 480
Qy 1190 ctgattgactgctgactgagacggcgagctgagtcgagcgagcgagccgctgagccaac 1249
Db 481 CTGATGATCTGCTGTAACGTGAACGCGCGCGCTGATGCGCGCGCGCGCGCGCTGCGCAC 540
Qy 1250 gtggaactggccggggcgaccccgagagctgtagtgcgactgagactgagactatcac 1309
Db 541 GTGGAACCTGGCCCGGCGCACCTGGAAGCTGTGATGCGGACTGGGACCTGGAATTTACATC 600
Qy 1310 gctacagcgagcaagcagcccaacacgctggtgagcagctgagactgaagccttacc 1369
Db 601 GCTTACCGCGGACGACGACGCGCCACACAGTGTGAGAGGAGCTGAGCTGGAAGCGCTTCAATC 660
Qy 1370 gacgagcgagtgccggcggtctaacatccggccggagtgatctgagtcggtgagagc 1429
Db 661 GACGAGCGGCTGCGCCGCGCTACATCCGCGGAGTGTATGATCGGTGAGAGC 720
Qy 1430 ggcctcgaactctggaagggcgggcgctctgagagagcgagcttctcgtlaagtgctg 1489
Db 721 GCGTTGGAACCTGTGGAGGGGCGGCGGCTGTGCGAAGCGCGGATTTTCCGTAACGCTG 780
Qy 1490 caagtag 1495
Db 781 CAGTAG 786

RESULT 4
LOCUS AF233376 1116 bp DNA linear BCT 02-AUG-2000
DEFINITION Streptomyces sp. 11A68 cellulase 12A (cel12a) gene, complete cds.
ACCESSION AF233376
VERSION AF233376.1 GI:9651812
KEYWORDS
SOURCE Streptomyces sp. 11A68.
ORGANISM Streptomyces sp. 11A68
REFERENCE 1 (bases 1 to 1116)
AUTHORS van Solingen, P., Melijer, D., van der Kleij, W.A.H., Barnett, C.C., Bolle, R., Power, S.D. and Jones, B.E.
TITLE Cloning and expression of an endocellulase gene from a novel
JOURNAL Streptomyces isolated from an East African soda lake
    Unpublished
    2 (bases 1 to 1116)
AUTHORS van Solingen, P., Melijer, D., van der Kleij, W.A.H., Barnett, C.C., Bolle, R., Power, S.D. and Jones, B.E.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2000) Microbial & Molecular Screening, Genencor International B.V., Archimedesweg 30, 2333 CN Leiden, The Netherlands
FEATURES
    Location/Qualifiers
```

BASE COUNT	ORIGIN	misc_feature	misc_feature	misc_feature	gene	misc_feature	CDS	source					
191 a	432 c	340 g	153 t					1. .1116 /organism="Streptomyces sp. 11AG8" /strain="11AG8" /db_xref="taxon:133452" 1. .1116 /gene="cell12A" 1. .93 /gene="cell12A" /note="Region: signal sequence" 1. .1116 /gene="cell12A" /note="endocellulase" /codon_start=1 /transl_table=1 /product="cellulase 12A" /protein_id="AA01283.1" /db_xref="GI:9651813" /translation="MRSHPRSATRTVLYVLASGALLTAAPAOANOICDRGCTTII ODRVLVNMRNGTSATOCINVTGGEFTTQDGSVPINGAPEKSTPSYDCHGNCAP RTLLPMRISISGSAPSSVSRYTNGVNYAAYDIDMDPTPTNGVNTETIMENRGA PVQDIPGSPVGAHVGSGEWMVNGSNDVIISFLAPSALISMSFDYKDEVDQAVSHG LATADWYLTSLQAGFEPMEGGTGLAVNSFSSAVNAGGNGCTPCTPAACOVSYTHW PGGTVDYTTITNGSTPEVDGWEIDFETLPAGHYTVSVNNALISPSASCAVFTARSTGNSNR IAANGTOSFGFOGTSSGAGTAPAGARLNGTCTCTA"					

Query Match	Similarity	7.0%	Score 131.8	DB 1	Length 1116
Best Local	Similarity	50.7%	Pred. No. 5,6e-11		
Matches 347	Conservative	0	Mismatches 312	Indels	Gaps
OY	ctgaagcccgagcgagacgtcagatcgtgtggagcgcgtggaagccgcgcatgtgtgcccggg	866			
Db	77 cggctccggcccgagcgaacacagcagatgtggacaccgtcaccggcacaccacagatccaa	136			
OY	ggcgctacccggtgtgacaaacgtaatgggctggcggaagacccgcagtcattgaatgtc	926			
Db	137 accggtacgtgtggtgcagaaacacccgtggggcaccagccaccacagtcgcatcatgtga	196			
OY	gactgtgaaacgggcgaacttcaagatcaacacggcgccgatatcgaacaaaggcaacacgtg	986			
Db	197 ccggcAACGGTTTCAGATCACCCAGGCGGACGAGTTCGGTGGCCGACCAAGGCGCCCA	256			
OY	ccgcctaccggcgaactactcgtggtgacgttgcacgttggggcgctgcgaacgaaattcgagt	1046			
Db	257 agtcttatccctcgcggtcttgcagacggctgcctactacggcgaacttgcggccgccacacacc	316			
OY	tgccgacggcgctgacgaaggatctgtccgaagtgtgcacgaagctgtgaacgtcaccagccatca	1106			
Db	317 tgcccatggcggtacagctgagtgaggagggcccccagcagtgctmctcttaccggtacacgg	376			
OY	cgaagcgcccgctgtgaatgcgcgcctacgaacatctgtattcaagtcgcgtacgaattccgca	1166			
Db	377 gcAACGGCGTCTTACACCGACCGGATCATGATTTGGTTCAACCGGGGTGGCCCGGATCCAC	430			
OY	acggtacacagcgcgcgccgcgcgcgtcgtatgtatgtcgtgaactgtgaacgagcgcggtatgc	1226			
Db	431 cCAAGGGGgTGAACCGGACCGGACCGATCATGATTTGGTTCAACCGGGGTGGCCCGGATCCAC	490			
OY	cgggagggacgacggctgtgacacacgtgtgaactgtccggggcgcaactgtggaaatctgtatg	1286			
Db	491 cCATGCGTTTCCGCGTCCGACGCGCCACAGTCGGGGCCGCGCGACGCTGAGGAGTGTGACCG	550			

OY	1287	cgaactggagactggaataatacatcgccttcacccgagcagacgacccacgaagtgcgtgagagc	1346
Db	551	GCAGCAACGCTTCGACACGACGTGATCTCTTCCTGCGCCCTCCCGGATCAGCACTGGAA	610
OY	1347	agcttgacctgaagagcccttcacatcgacagcagcgagtcgcccgcggtacatccgcggagtc	1406
Db	611	GCTTGACGATCAAGGACTTCGTCGACACAGGCCGCTGACGCCGCTGGCCACCCCGACT	670
OY	1407	ggtatctgcagtcggtgtagaagcgggcttcgaactcttggaagggcggggcggtctgcgaa	1466
Db	671	GCTTACTCTACCAACGATCCAGCGCGGCTTCGAGCCGTGGAGGGGCGCACACGGTCTGGCCG	730
OY	1467	gcgcgcattttccgtlaacgyltga	1491
Db	731	TGAAGCTCGTTCCTCCGCCGGTGAA	755
RESULT	5		
LOCUS	ARI29926	1116 bp	DNA
DEFINITION	Sequence 2 from patent US 6187577.	linear	PAT 16-MAY-2001
ACCESSION	ARI29926		
VERSION	ARI29926.1	GI:14117823	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1116)		
AUTHORS	Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Weyler,W.		
TITLE	Cellulase producing Actinomycetes cellulase produced theferom and method of producing same		
JOURNAL	Patent: US 6187577-A 2 13-FEB-2001;		
FEATURES	Location/Qualifiers		
source	1..1116		
	/organism="unknown"		
BASE COUNT	191 a 432 c 340 g	153 t	
ORIGIN			

Query Match	Similarity	Score	DB	Length
Best Local	50.7%	Pred No. 5.6e-11;		
Matches 347;	Conservative	0;	Mismatches 332;	Indels 6; Gaps 1.
QY	807	ctgaagcccgagcgcagcctcgaactcgtctgcgaagcgtctggagacgcgcgcgaatgtctgcccggg	866	
Db	77	cggctccggcccaaggcgaaccagagatctggaccgcgctacagggcaaccacagatccag	136	
QY	867	gagcgtaccgggtgatcaacaacgatatgggcgcggagagacgcgcccagttgattgagttgc	926	
Db	137	accggttagctggtgtcaggaacacacccctgggggaccacagcccccacagtcataatgtga	196	
QY	927	gacttgaaacgcgcacacttcacgatccacacgcgcgcgcgatcaagacaagggcaacaacgttg	986	
Db	137	ccggcacacgctttccagatccaccacagcccgacgcttccgtgccgcacacacagccgcccga	256	
QY	987	ccgcatactcgcgcacatactactcttcgtggctgcacatgggcgcctctcagacgaattcgggat	1046	
Db	257	agtcctaatccctcgggtctacgacagcgctgcactaacgaaactggcgcccccacgacagcc	316	
QY	1047	tgcgcgcgcgcgttcagagaagctgtctcgaacgttcgcacgaactcgaagctcgaacccgata	1106	
Db	317	tgcccatcgcggatcagctcgcgctgacggcagccgccacagctgcttcctacacacgc	376	
QY	1107	cgaagcggcgcgcgtgaaatgcgcgcatacgaatctgtttcaagctccgttcaagaattccgca	1166	
Db	377	gcacggcgctctacacacgcgcgctacgacatctgg-----ctggacccgcgaaccccgga	430	
QY	1167	acgcgtacagcgcgcgcgcgcgcgtgatctatctgtctgaactctgaagacgcgcgcgtgatgc	1226	
Db	431	ccaaagggggtgaaccggacggagatcattgattggtttcaacggggctggcccggtccagc	490	
QY	1227	cggcgcgacgcgcgcgttgaccacgttgaacttgcgcgggcgcacactcgggaagtctgattg	1286	

Db	491	CCATGCGTTGGCGCGTTCGGACAGGCCCAAGCTGTGGGGCGGCAGCTGTGGAGGTGTGTGACC	550
Qy	1287	cgcacttgtagctgtgaattacatcgcctccacgagcgcaagaagccaccacgltcgtgtgaagcg	1346
Db	551	GCAGCMAAGGTTTGCAAGCACACAGTGANTCTCTTCCTGGCCCCCTCCCGATTCACACACTTGA	610
Qy	1347	agcttgacctgaaggaccttcatacgcagcagcgagtcggccgcgtgtacatlcsgccgagagt	1406
Db	611	GCTTTCGACGCTCAAGGACTCTTCGTGCACAGGCGCGTGCACGACCGCTGGCCACCCGGA	670
Qy	1407	ggtatctgcatagcggttggaaagcagcgcttcgaactcctcgtgaagggcgagggccgagtlcgcga	1466
Db	671	GSTACTCATCCACGAGTTCACGGCGGCGCTTCGAGACCCTGGAGGGCGGCAACCGGTCTGGCGC	730
Qy	1467	gcgcgcgaatlcttcggtlaacggtlga	1491
Db	731	TGAACCTCGTTCTCTCCYCCGCGGTGA	755
RESULT	6		
LOCUS	ARI68360	1116 bp	DNA
DEFINITION	Sequence 2 from patent US 6287839.	Linear	PAT 17-DEC-2001
ACCESSION	ARI68360		
VERSION	ARI68360.1	GI:17904235	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1116)		
TITLE	Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Weyler,W.		
JOURNAL	Cellulase producing actinomycetes, cellulase produced therefrom and		
FEATURES	method of producing same		
source	Patent: US 6287839-A 2 11-SEP-2001;		
	Location/Qualifiers		
	1..1116		
	/organism="unknown"		
BASE COUNT	191 a 432 c 340 g	153 t	
ORIGIN			

Query Match	7.0%	Score 131.8	DB 6	Length 1116
Best Local Similarity	50.7%	Pred. No. 5.6e-11		
Matches 347	Conservative 0	Mismatches 332	Indels 6	Gaps
QY 807	ctgagcccgagccgagccgctcgagctctgctgagagctctggagccgctgcgagctgtgctggcgggg			
Db 77	cggcttcggcccgccagcgaaaccagagatctcgagccgctacgagccgaaaccagagatccaggg			
QY 867	ggcgctaccgggtgtgtatccaacaacgtatctggcgcgggagagacggccgagtcatgtggctcg			
Db 137	accggctacgctggctggcagaaacacccctggggcgaccacggccaccacagtcacatcaattgtca			
QY 927	gactgaaacggcgcaacttcagatcagatcacagcgcgccgcatcagcacaagcgcaacaacgctgg			
Db 197	ccggcagacgggtttcgagatcacaccagggccgagaggttcgctggccgacaaacggcgcccgca			
QY 987	ccgctatccggcgccatctactctctgggtgtgccactgtggcgcgctctgcaagagcaatctcggaat			
Db 257	agttcttttccctccgctttacgacggcctccacatagcggaacgtgcggcccccaccagacgcg			
QY 1047	tgccgagcgcgctgagagagctgtctcgagcgtgtgcgaagagctgagagcgtctcaagccgatca			
Db 317	tgcccatcgggatcagctacgtctcgatggcagcggcccccagcagtgcttcttcaaccgcttacaccg			
QY 1107	cgacggcgcgctgagaaatgacgctacgacatcttgctcaagtcaccgctcaagaaatctccgca			
Db 377	gcacagcgcggtctacacaccccgctacgacatctgg-----ctgacaccgacaccccgca			
QY 1167	acggttaacagcgcgcgcgccgagctcgatcttgctgaactctgaaacggcgcgagctgcatgct			
Db 431	ccaacggcgagtgaaaccgagaccgagatcatgattctgtttcaaacggcggtggcccccgggtccacg			

OY	1227	cggggagcgaacccgcgttggcccaccgctggagacttgccggagcccaacctgggaaagtctgtagatg	1286
Db	491	CCATCGGTTCCGCCGTCCGCACGGCCCCACAGTGGCGGCGCCGACACTGGAGGTGTGGACCG	550
OY	1287	ccgaacttgagacttgatattacatcgcctctaaccggtcgacgagcacccaacacgttcgttgagcg	1346
Db	551	GCAGCAGACGGTTTCGAACGACGTGATCTCTTCTTGCGCGCCCTCCCGATTCACAGCACTGGA	610
OY	1347	agcttgagaccttgagaaggcctcatcgaaacgacgcygttgcgccgcggtcttacatctcgccggaggt	1406
Db	611	GCTTCGACGCTCAAGGACACTTCGTGCGACAGGCGCGTCAGCGCACGCGCTGGCCACCCTCGGACT	670
OY	1407	ggtatctgatcgatcggttggagagggccttcgaaactcttggaagggcgggccggtctctgcgaa	1466
Db	671	GGTACTCATCCACGACATCCAGCGCGGCTTTCGAGCCGTGGAGGGCGGCGACACCGGTTGCGCG	730
OY	1467	gcgcgcgaattttccglaaacgctga	1491
Db	731	TGAAGTCGTTCTCTCCGCGGGTGA	755
RESULT	7		
LOCUS	AR168362		
DEFINITION	Sequence 4 from patent US 6287839.	DNA	linear PAT 17-DEC-2001
VERSION	AR168362		
ACCESSION	AR168362.1		
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1716)		
AUTHORS	Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Weyler,W.		
TITLE	Cellulase producing actinomycetes, cellulase produced therefrom and method of producing same		
JOURNAL	Patent: US 6287839-A 4 11-SEP-2001;		
FEATURES	Location/Qualifiers		
source	1..1716		
BASE COUNT	307 a 611 c 559 g	239 t	
ORIGIN	/organism="unknown"		

[illegible]

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	TITLE	JOURNAL	REFERENCE	AUTHORS	FEATURES	gene	CDS	BASE COUNT	ORIGIN
Db	1065	TCGAGCTCAAGGACTCTTCGACCAAGGCGCTTCGACGCCAGGCGCTGCGCCACCCGGACTTGG	AF145250	1944 bp	DNA	linear	BCT 09-JUN-1999								385 a	559 c 648 g 352 t
QY	1409	tatctgcgatcggtgtgaagcaagcggtcttgcaactctgtggagggcgagggcggtctgcggaagc	AF145250	1944 bp	DNA	linear	BCT 09-JUN-1999								385 a	559 c 648 g 352 t
Db	1125	TACCTCACCACCATCCAGCGGGCTTTTCGACGCGTGGAGGCGGCGGACACCGCTTCGGCCGCTGG	AF145250	1944 bp	DNA	linear	BCT 09-JUN-1999								385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
QY	1469	gccgattctcccgtaacggtgca	AF145250.1	GI:5020003											385 a	559 c 648 g 352 t
Db	1185	AACCTGCTTCTCTCCGCGGCTGMA	AF145250.1	GI:5020003												

Dn	1541	AGAACTTATCCGGAAAAACAAGG-CAGATTCCGGTAAGTCGTCTTTGCCTCGCATCTCA	1599
OY	1708	gaagcagggatttcttagtaaacccacgaaagcgctccgcgaagtggcctcaaccgcgcagc	1767
Dn	1600	GAAACACTGTTTGTAGTAAGCACGACGAAGGCCGCTGACGTAOTCS-GTTAACCCGGACGG	1658
OY	1768	cacttcgtctgcctccgctgcctgaactaagagcgcttgccgctgtttcaaacggcagct	1827
Dn	1659	CACCTCCTGCTGGTCCGCGCCCTGACTTAAAGAAGCAGTGGGCTGTCTTCAAACCTCAGCG	1718
OY	1828	ataagcatcacagctcgaatcgcgaactctccgac	1865
Dn	1719	AACCGTATGTA-AGGTCCAATTCGCAATTCCTCTTGAC	1755
RESULT	9		
LOCUS	SHU51222	3527 bp DNA linear BCF 23-JUN-1997	
DEFINITION	Streptomyces halstedii beta-1,4-endoglucanase (celA2) gene, complete cds, and cellulose binding protein (p40) gene, complete cds.		
ACCESSION	U51222		
VERSION	U51222.1	GI:2209259	
KEYWORDS	.		
SOURCE	Streptomyces halstedii.		
ORGANISM	Bacteria; Filicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces. Garda,A.I., Fernandez-Abalos,J.M., Sanchez,P., Ruiz-Arribas,A. and Santamaria,R.I. Two genes encoding an endoglucanase and a cellulose-binding protein are clustered and co-regulated by a TPA codon in Streptomyces halstedii JMB Biochem. J. 324 (Pt 2), 403-411 (1997)		
REFERENCE	JOURNAL	97307849	
MEDLINE	2 (bases 1 to 3527)		
AUTHORS	Santamaria-Sanchez,R.I.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-MAR-1996) Ramon I. Santamaria-Sanchez, Instituto de Microbiologia Bioquimica, Consejo Superior de Investigaciones Cientificas / Universidad de Salamanca, Avda. del Campo Charro, s/n, Salamanca, Salamanca 37007, Spain Location/Qualifiers .3527		
FEATURES	source	/organism="Streptomyces halstedii" /strain="JMB"	
	gene	/db_xref="taxon:1944" 595..1728 /gene="celA2" 595..1728 /gene="celA2" /EC_number="3.2.1.4" /note="beta-1,4-endoglucanase" /codon_start=1 /transl_table=1 /product="Cel2" /protein_id="AAC45429.1"	
	CDS	/db_xref="GI:2209260" /translation="MRALPHGRPRPSRLSLGLLALATVAAYLAAPRAADTLVCEG YGSITIGRRVVQNNRWGASAPQCVTATDSCFRTYQAQAVPTNGAPKSYSPVNGCH VTSNPGTNPDAQVSGIASAPSISLYGVGSAVYNASVDILDPKNGVNRTIMIM WLNKGVLPIPIGSOAGSVGRPMWRGNSNDYISFYAPSAVSAMDVPDI NTIARGMAONMYLTLSVOAGEPMPONGAGIAVNSFPSSVNLGFTGGSFGA RGSPVRA CEVAYATRNWSDEGTADVANTSTPYDSRRLSFLLPSGGRVTRNANATVSGTPGAN TATGLAHSAOVAAASQTFEGQTYSGFESKPSGSLMGARCA" 1945..3039 /note="cellulose binding protein" /codon_start=1 /transl_table=1 /product="p40" /protein_id="AAC45430.1"	
	CDS		



Db 554 TCGTGACAAAGCCGTCGTACAAAGCGCTGGTACGATCATCTGGCT-----GGACCCGACGC 607

Qy 1161 cgggaacgsgcttaacgcgsgcgsgcgsgctgatatcttgctgtgaactggaacgsgcg 1220

Db 608 CCCGACCCGACGGGGTGAACCCGACCGAGATCATATCTGGTTCAACCGGGTGGGGCAGA 667

Qy 1221 tgatgcgsgcgcgacgcgcgctgctccacgctgaaacttggaactgcgsgggcacaactggaagtc 1280

Db 668 TCCACCCGATCGGCTCGCAGGTGGGACCGCGCTCCGTGGCCGGGGGCGACCTGGAGGTGT 727

Qy 1281 ggtatgcggaactgsgaactggaattacatcgcctaacgcgsgcaagccccacacgctcg 1340

Db 728 GGTCCGGGGGGAACGGCACCAACGAGCTGCTCTCTGCTCCCTCGTCCGGGATGAGCA 787

Qy 1341 tgaagcagcttgaccttgaagagccttcacatcgaaagcsgctgcgcgsgctacatccgac 1400

Db 788 GCTGGAGCTTCGACCTGATGGACTTCGTCTCGGGCCACCGTCGCCCGGGGGCTGGGGGCA 847

Qy 1401 cggagtgatctcgcacgctgaaagcgctgaaagcggctcgaaacttggaaagcgsggscgctc 1460

Db 848 ACGACTGGTACTCGACGAGCATCCAGCGCCGGCTTTGGAGCCCTCGCAAGAACGGTGGGGTC 907

Qy 1461 tgcgaagcgccgattttccgtaaaagcggtgacgtacgttcgacacgaagcaccag 1514

Db 908 TCGCGGTAACTCTTCTCTGTCGACCGGTAACACGGCGGCTCGGACGAACCTTG 961











Accession	Sequence	Position
Oy	gcgacgagcccccacgcgtctgctgagcagctcgagccctgaagcctctatctgaagcagac	1377
Oy	1318	
Db	CGAGCTCACGTACCGGGCGAGGTCCGCCACAGAGACGTCCGACTGTACCGCGCGCCG	7657
Db	7598	
Oy	ggtctgcgcgcggtctactctcgcgcgcgaagtggtgtctgcgtatgctggtgagagcgtctga	1437
Oy	1378	
Db	CCACACAGGCGCGCTTCCCGCGCCCTCGAGTGCGCAGGTCAACGGCGGTACGCCCCCGGGT	7717
Db	7658	
Oy	actctggaagggcgggagcgcgtctg	1462
Oy	1438	
Db	GGCCGAGAGACCGCGATCCCGCTG	7742
Db	7718	

RESULT	15
AF335723	
LOCUS	AF335723 4262 bp DNA linear BCT 11-FEB-2001
DEFINITION	Burkholderia pseudomallei formate hydrogenlyase subunit 4 and hydrogenase-4 component B genes, complete cds.
ACCESSION	AF335723
VERSION	AF335723.1 GI:12744948
KEYWORDS	
SOURCE	Burkholderia pseudomallei.
ORGANISM	Burkholderia pseudomallei

CDS

CDS

BASE COUNT  
ORIGIN

Query Match	3.78;	Score 68.8;	DB 1;	length 4262;
Best Local Similarity	47.38;	Pred. No. 0.18;		
Matches 208;	Conservative	0;	Mismatches 232;	Indels 0;
			Gaps	0;

Oy	CGAGAGCCTGAGCCGAGACCGTCTGAGAGCTGTGGAGACGCTGGAGGCGCGGATATGG	860
Db	CGCAGCACGCGACAGAAACAGCATGAAGAGCCCGCATGTACCTCGCCGCGCGGTGG	1465
Oy	CCGAGGAGCGCTACCGGTTGATCAACAAGTATGGTGGCGAGAGACGCCCGCATGTCAATG	920
Db	ACACGGGCGCGAATAGCAGGCGACCCCGTATGTTGGTCCCGGATGCCCGCGCAGACAGCG	1522

Qy 921 agtgcgactgtgaacggcgcaacttcagatcacacgycgcgatcacgacaagcaca 980  
| | | | | | | | | | | | | | |  
Db 1524 cggccgtgaggcactcgacgaccttcacgtactgaacggcgccgagccgacgacnaccg 1583

Oy 981 acgtggccgcctatccgagcatctactcgggtgccaattggcgctgcaggcaatt 1044  
||| ||| | | | | | | | | | | |  
Db 1584 TCGGGCGCCGACCAAGCGGATCCGTCCCAAGGGCTTCGCATCGCGCGCGGT 1643

Oy 1041 cgggattgcgcgcgcgtcagagctgtccgacgtcgcaagctgaacctaacgc 1100  
||| ||| ||| | ||| ||| ||| ||| |||  
Db 1644 CGAGCTTGCCCTCGGGGCGGCCGACGAAACAGATGCCGAAGAACCTTGCATTCGCGT 1703

Dy 1101 cgatcagacgagccgtgaattgcgccctaacacatctgttcagtcgccgcgaatt 1166  
+ ||| | | | | | | | | |  
Db 1704 AGCCGGCAGCGCGCGCTCAGCGGACGAGGCCCGGCAGCACCGTATCCGA 1766

Dy 1161 ccgcaacgcgtacagcgcgccgaagtatgatctgtcgtaaatcggcgcg 1220  
| | | | | | | | | | | | | | | |  
Db 1764 GCACCGAATCCGCGAGCCCCGGCGTGACAGGAAGCTTCACAGAGCACCATTCGACA 1820

Qy	1221	tgatgcgcgycgcagccgc	1240
Db	1824	CGAGCCGCTCGACGGCGC	1843

Search completed: August 19, 2002, 12:44:21  
Job time: 10405 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 09:55:01 ; Search time 544 Seconds

(without alignments)  
5923.986 Million cell updates/sec

Title: US-10-003-759-1

Perfect score: 1877

Sequence: 1 gggcgagtcgaacgtcgggc.....tctccgacacttaaacac 1877

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq\_032802:\*  
1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.8	7.0	1116	20	AAV72359
2	131.8	7.0	1116	21	AAZ57029
3	131.6	7.0	1715	21	AAZ57031
4	57.6	3.1	1983	22	AA54513
5	53.6	2.9	77536	21	AAAI4651
6	51	2.7	58857	21	AA58471
7	49.8	2.7	756	17	AAAT16766
8	49.4	2.6	1404	21	AAA38763
9	49	2.6	756	14	AAO43032

10	48.8	2.6	985	19	AAV64548	M. tuberculosis im
11	48.8	2.6	985	19	AAV44439	Mycobacterium tube
12	48.8	2.6	985	20	AAZ19349	M. tuberculosis an
13	48.8	2.6	985	20	AAZ19137	M. tuberculosis re
14	48.4	2.6	12152	22	AA508699	M. carbonaceus DNA
15	48	2.6	4403765	22	AA199683	Mycobacterium tube
16	47.8	2.5	24379	18	AA93095	Streptomyces fireo
17	47.8	2.5	24379	19	AAV25925	Streptomyces roseo
18	47.6	2.5	30001	18	AAV61016	Total DNA sequence
19	47.6	2.5	30001	20	AAO5110	S. aureofaciens DN
20	47.6	2.5	114955	20	AAV53491	Human adenosine AI
21	47.2	2.5	1908	16	AAO99364	S. lividans protea
22	47.2	2.5	1908	20	AAV84065	Tripeptidyl aminop
23	47.2	2.5	1908	21	AAV61403	cDNA sequence enco
24	47.2	2.5	77536	21	AAAI4651	Nucleotide sequenc
25	46.8	2.5	2712	19	AAV94214	Polyangium brachys
26	46.8	2.5	15872	21	AAZ87283	S. venezuelae vep
27	46.8	2.5	20394	22	AAV24892	Plasmid biosynth
28	46.4	2.5	109519	22	AA508693	Micromonospora DNA
29	46	2.5	44377	18	AAV78508	Platenolide syntha
30	46	2.5	44377	18	AAV80414	Human adenosine AI
31	46	2.5	114955	20	AAV53491	Myceliophthora the
32	45.6	2.4	744	18	AAV74073	H. virescens inorg
33	45.6	2.4	867	22	AAH48828	S. avermitilis ave
34	45.4	2.4	31422	21	AAH92302	Streptomyces averm
35	45.4	2.4	31422	22	AAH79278	Mycobacterium tube
36	45.2	2.4	1291	20	AAH87940	M. tuberculosis im
37	45.2	2.4	2852	19	AAV64558	Mycobacterium tube
38	45.2	2.4	2852	19	AAV44449	M. tuberculosis an
39	45.2	2.4	2852	20	AAZ19359	M. tuberculosis re
40	45.2	2.4	2852	20	AAZ19147	Non-reducing sacch
41	44.8	2.4	2268	21	AAAI0501	Rice 1-deoxy-D-xy
42	44.8	2.4	2618	21	AAA38760	Non-reducing sacch
43	44.8	2.4	3252	21	AAAI0505	Mouse Sox1 cDNA.
44	44.6	2.4	2376	20	AAV16153	L05390 cDNA clone.
45	44.6	2.4	5224	22	AAV90079	

## ALIGNMENTS

RESULT 1	
AAV72359	standard; DNA; 1116 BP.
AAV72359	
AAV72359	
28-JUL-1999	(first entry)
Actinomyces sp. 36KD cellulase DNA.	
Cellulase; detergent; animal feed; nutritional value; textile;	
stone washing; texture modification; appearance; cellulosic fabric;	
pulp; draining; paper; baking additive; starch treatment; grain;	
high-fructose corn syrup production; ethanol production; fibre reduction;	
milling; ss.	
Actinomyces sp.	
WO9925846-A2.	
27-MAY-1999.	
18-NOV-1998;	98WO-US24649.
24-JUN-1998;	98US-0104308.
19-NOV-1997;	97US-0974041.
19-NOV-1997;	97US-0974042.
(GENV ) GENENCOR INT INC.	
Jones BE, Van Der Kleij WAH, Van Sollingen P, Weyler W;	









```
FT CDS complement (12113..13150)
FT /tag= l
FT /note= "fkdb gene"
FT complement (13212..23988)
FT /tag= m
FT /note= "fkdb gene"
FT complement (13452..13662)
FT /tag= n
FT /note= "ACP6"
FT complement (13761..14394)
FT /tag= o
FT /note= "KR6"
FT complement (14517..15294)
FT /tag= p
FT /note= "ER6"
FT complement (15438..16587)
FT /tag= q
FT /note= "dehydratase domain (DH) 6"
FT complement (16587..17820)
FT /tag= r
FT /note= "acyltransferase domain (AT) 6"
FT complement (17820..19053)
FT /tag= s
FT /note= "KS6"
FT complement (19116..19326)
FT /tag= t
FT /note= "ACP5"
FT complement (19464..20097)
FT /tag= u
FT /note= "KR5"
FT complement (20241..21420)
FT /tag= v
FT /note= "DH5"
FT complement (21420..22653)
FT /tag= w
FT /note= "AT5"
FT complement (22653..23892)
FT /tag= x
FT /note= "KS5"
FT complement (23992..46573)
FT /tag= y
FT /note= "fkdb gene"
FT complement (24163..24373)
FT /tag= z
FT /note= "ACP4"
FT complement (24997..26146)
FT /tag= aa
FT /note= "DH4 (inactive)"
FT complement (26146..27430)
FT /tag= ab
FT /note= "AT4"
FT complement (27430..28684)
FT /tag= ac
FT /note= "KS4"
FT complement (28750..28960)
FT /tag= ad
FT /note= "ACP3"
FT complement (29092..29740)
FT /tag= ae
FT /note= "KR3"
FT complement (29869..31018)
FT /tag= af
FT /note= "DH3 (inactive)"
FT complement (31018..32185)
FT /tag= ag
FT /note= "AT3"
FT complement (32185..33439)
FT /tag= ah
FT /note= "KS3"
FT complement (33505..33715)
FT /tag= ai
FT /note= "ACP2"
FT complement (33823..34480)

FT /tag= aj
FT /note= "KR2"
FT complement (34606..35749)
FT /tag= ak
FT complement (28750..28960)
FT /tag= al
FT /note= "DH2 (inactive)"
FT complement (35749..37144)
FT /tag= am
FT /note= "AT2"
FT complement (37145..38296)
FT /tag= an
FT /note= "KS2"
FT complement (38371..38581)
FT /tag= ao
FT /note= "ACP1"
FT complement (38677..39307)
FT /tag= ap
FT /note= "KR1"
FT complement (39442..40609)
FT /tag= aq
FT /note= "DH1"
FT complement (40609..41842)
FT /tag= ar
FT /note= "AT1"
FT complement (41842..43093)
FT /tag= as
FT /note= "KS of extender module 1 (KS1)"
FT complement (43144..43660)
FT /tag= at
FT /note= "ACP of loading domain"
FT complement (43777..44629)
FT /tag= au
FT /note= "ER of loading domain"
FT complement (44974..46573)
FT /tag= av
FT /note= "CoA ligase of loading domain"
FT 46754..47788
FT /tag= aw
FT /note= "fkdb gene"
FT 47785..52272
FT /tag= ax
FT /note= "fkbp gene"
FT 52275..71465
FT /tag= ay
FT /note= "fkba gene"
FT 52362..53576
FT /tag= az
FT /note= "KS7"
FT 53577..54716
FT /tag= db
FT /note= "AT7"
FT 54717..55871
FT /tag= ba
FT /note= "DH7"
FT 56019..56819
FT /tag= bc
FT /note= "ER7"
FT 56943..57575
FT /tag= bd
FT /note= "KR7"
FT 57710..57920
FT /tag= be
FT /note= "ACP7"
FT 57990..59243
FT /tag= bf
FT /note= "KS8"
FT 59244..60398
FT /tag= bg
FT /note= "AT8"
FT 60399..61412
FT /tag= bh
FT /note= "DH8 (inactive)"
```





DB 578 gactctgcaagccgcaagctgctgctggaacggctgtgtccaccgggtgtctccggagccctg 637  
 QY 1252 ggaactggccgggagccactgtggaactgtgtatgccgactgtggaactgaattaccctgc 1311  
 DB 638 cgggcccgcgcaagctgtgctgtgacccgctggtgtccaccgggtgtctccggagctgcagggc 697  
 QY 1312 ctaccggcgacagcagccaccacagctcgg 1340  
 DB 698 cgcacaggtgtctcagcagccacagcagctccg 726

RESULT 8  
 AAA38763  
 ID AAA38763 standard; cDNA: 1404 BP.  
 AC AAA38763;  
 XX  
 DT 14-SEP-2000 (first entry)  
 DE Wheat 1-deoxy-D-xylulose 5-phosphate synthase putative coding sequence.  
 XX  
 KW Wheat: 1-deoxy-D-xylulose 5-phosphate synthase; DXPS;  
 XX Isoprenoid biosynthesis; herbicide: ss.  
 OS Triticum aestivum.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 31..1404  
 FT /tag= a  
 FT /product= "1-deoxy-D-xylulose 5-phosphate synthase"  
 FT /partial  
 XX  
 PN WO200032792-A2.  
 PD 08-JUN-2000.  
 XX  
 PF 02-DEC-1999; 99WO-US28587.  
 XX  
 PR 03-DEC-1998; 98US-0110779.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Cahoon RE, Tao Y, Williams ME, Coughlan SJ, Weng Z;  
 XX  
 DR WPI: 2000-412338/35.  
 DR P-PSDB: AAY97425.  
 XX  
 PS Claim 2; Page 67; 73pp; English.  
 XX  
 CC The present sequence is a putative coding sequence for the wheat  
 CC 1-deoxy-D-xylulose 5-phosphate synthase enzyme (DXPS). Its protein is  
 CC involved in the isoprenoid biosynthesis pathway. The cDNA was identified  
 CC by sequencing a number of clones and then comparing their protein  
 CC sequences to known proteins; this showed the sequence's similarity to the  
 CC Capsicum annuum DXPS sequence. The DXPS gene and protein can be used to  
 CC create transgenic plants which express the gene at either different  
 CC levels or at different stages of development compared to normal, and to  
 CC identify herbicides.  
 CC  
 XX  
 SQ Sequence 1404 BP; 277 A; 482 C; 439 G; 206 T; 0 other;

Query Match 2 6%; Score 49.4; DB 21; Length 1404;  
 Best Local Similarity 46.9%; Pred. No. 0.13;  
 Matches 220; Conservative 0; Mismatches 246; Indels 3; Gaps 2;  
 QY 938 ggaacttcaacgacgagcgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 997  
 DB 355 ggcgtctgtcagctcagcgcgtcgcgtcgtcactacgtcttcaacaccccgacgacgacgacg 414

QY 998 gccatctactcgggtgccaacttgaggcgctcgcagcagcaatttcggatctgcgcggcgc 1057  
 DB 415 cctcttggaagctcggccaccagctcgtatcccgacacaaagattctcagggggcgggcgataag 474  
 QY 1058 gtgcaggaactgtctgcagcgtgcgcagcagcttggaacgtcaccgcgacatcaagacggcgc 1117  
 DB 475 atgcgaacgatg--cgcagacacaaagcctgtccggtctgtcgaagcgtcccgag-aagc 531  
 QY 1118 tggaaatgcgcctcagacatctggttcaatctcccgtaacgaatttcggacagcgctacagc 1177  
 DB 532 gaggtaacagagcttcggcccgccagccacagctccacacacattccgcgcgcctcggaatg 591  
 QY 1178 ggcggcgccgagctgaatgtatctggtcgaacttgaaacggcggtgtgacgtcgggcgccagc 1237  
 DB 592 gccgtcgggagaggaactcaaggcgcggaagacaacagtggtggtggtgtgttggtggagcggg 651  
 QY 1238 cgcgtgccaacgcttggaactggtgcggggccacacctgggaagctgtgtatctgcgacttggaac 1297  
 DB 652 gccatgacgagcgcgagggcgtacagagcgatgaaacaacggccggtaccctcagactcgagac 711  
 QY 1298 tggaaattacatcgctaccgcgccagcagccacacagctcggtgtgagcgacttggaacctg 1357  
 DB 712 atgattcgtatctcctcaacgacaacaaagcaggtgtgtcgtccgacggcgacgtcgaagg 771  
 QY 1358 aaggccttcatcgcagcagcgggtcgcccggtctacatccggccggagt 1406  
 DB 772 ccggcgccggccggtggggcggtcagcggtccctcagcaagctgtgagt 820

RESULT 9  
 AAQ43032  
 ID AAQ43032 standard; DNA: 756 BP.  
 XX  
 AC AAQ43032;  
 XX  
 DT 07-SEP-1993 (first entry)  
 XX  
 DE Collagen-like polymer DCP3-C2(AB12)C2 coding sequence.  
 XX  
 KW Recombinant: collagen-like polymer; CLP; tripeptide; helix; membrane;  
 KW fibre; film; coating; triad sequence; collagen; mammalian; moulding;  
 KW hydrogel; interchain linkage; colloid suspension; DCP; antibody; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9310154-A.  
 PN  
 XX  
 PD 27-MAY-1993.  
 XX  
 PF 04-NOV-1992; 92WO-US09485.  
 XX  
 PR 12-NOV-1991; 91US-0791960.  
 XX  
 PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.  
 XX  
 PI Cappello J, Ferrari FA;  
 XX  
 DR WPI: 1993-182496/22.  
 DR P-PSDB: AAR37738.  
 XX  
 PT High mol. wt. collagen-like protein polymers - capable of being  
 PT produced in unicellular microorganisms  
 XX  
 PS Disclosure; Page 39; 82pp; English.  
 XX  
 CC The sequences given in AAQ43030-34 encode examples of recombinantly  
 CC produced DCP collagen-like polymers (CLPs) which consist of repeated  
 CC tripeptide sequences selected from a wide range of GXY sequences,  
 CC where X and Y can be any amino acid. These sequences can be cloned  
 CC into plasmids and used to transform E. coli to produce the DCP  
 CC proteins. DCP peptides comprise repeated units of: A = GAPAPGPP,  
 CC B = GSRGDREPP and/or C = GAHGRAGPK. These polymers may be used to



ID AAV4439 standard; DNA; 985 BP.  
 AC AAV4439;  
 XX  
 DT 09-NOV-1998 (first entry)  
 DE Mycobacterium tuberculosis antigen XP25 3' DNA.  
 XX  
 KM Tuberculosis; infection; diagnosis; antigen; XP25; ss.  
 XX  
 OS Mycobacterium tuberculosis strain Erdman.  
 XX  
 PN W09816645-A2.  
 PD 23-APR-1998.  
 XX  
 PF 07-OCT-1997; 97WO-US18214.  
 XX  
 PR 13-MAR-1997; 97US-0818111.  
 PR 11-OCT-1996; 96US-0729622.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1998-251292/22.  
 XX  
 PT New isolated Mycobacterium tuberculosis polyepitides and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and diagnosis of tuberculosis  
 XX  
 PS Claim 4; Page 187-188; 250pp; English.  
 XX  
 CC This is the 3' region of DNA coding for an antigenic portion of  
 CC Mycobacterium tuberculosis antigen XP25; 5' DNA is provided in  
 CC AAV4439. XP25 DNA was isolated from a M. tuberculosis strain Erdman  
 CC genomic DNA expression library using sera from patients having  
 CC extrapulmonary tuberculosis. It bears no similarity to known  
 CC sequences. The invention relates to methods and compositions for  
 CC diagnosing tuberculosis. It provides polypeptides (see  
 CC AAW64291-W64379) comprising an antigenic portion of a soluble M.  
 CC tuberculosis antigen, or an immunogenic portion of a M.  
 CC tuberculosis antigen, as well as DNA sequences encoding such  
 CC polypeptides, recombinant expression vectors and transformed or  
 CC transfected host cells. Also claimed are methods and diagnostic  
 CC kits for detecting M. tuberculosis infection in a patient using  
 CC these polypeptides, antibodies or oligonucleotide probes and  
 CC primers.  
 CC  
 SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 2.6%; Score 48.8; DB 19; Length 985;  
 Best Local Similarity 44.9%; Pred. No. 0.16;  
 Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

QY 773 ttcccgatgagcacaacggaagaccgagcctgagcccgagccgctgagcgtg 832  
 II III IIII  
 DB 175 ttccgagtgagcgccgagccagacagtgagcgccgagccgagccgagcg 234  
 QY 833 tgcgagcgttgagacgagcgagctgagcgagggcgctaccgggtgatcaacgta 852  
 III  
 DB 235 accggtaccgagcgagccgagcggtgtgagcgagccagcggtgtaagacatcgcg 294  
 QY 893 tgggagcgagagacgcccagtgatgaggtgagcgtgtaacgagcaactcaagatc 952  
 IIII  
 DB 295 gccgagcgagcgagcggtgtaagcgagctgggagccagcggtctggcctctcc 354  
 QY 953 acacgagccgatcaacagcaacgagtgagccgcatccgagcactactctcgag 1012  
 I  
 DB 355 ggcttgaagcgagcgagcaagcgagcgagcgagcgagcgagcgagcgagcgagc 414

QY 1013 tgcacatgggagcgagcgtgcaagc-----aattcgagattgagcgagcgagtgcaaggc 1068  
 I  
 DB 415 atcaacggggcgagcgagccgagccgagcaacgagcgagcgagcgagcgagcgag 474  
 QY 1069 gtccgagcgtgagcagcagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 1128  
 I  
 DB 475 gccgagcgttcgagcacaacgagcgaggtcgagcgagcgagcgagcgagcgag 533  
 QY 1129 ctacgacatcgttgcagtcacgaaatccgagcaacgagcgagcgagcgagcgag 1188  
 I  
 DB 534 cggcaacgagcgagcaacgagcgagcgagcgagcgagcgagcgagcgagcgagcg 593  
 QY 1189 gctgatgatcgtgagcagcagcgagcgagcgagcgagcgagcgagcgagcgag 1248  
 DB 594 gggcaatgagcgagcaacgagcgagcgagcgagcgagcgagcgagcgagcgagcg 653  
 QY 1249 cgtgagacgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1308  
 DB 654 cggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 713  
 QY 1309 cgcctacgagcgagcagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1362  
 DB 714 cactgatcgagcgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 767

RESULT 12  
 AA219349  
 ID AA219349 standard; DNA; 985 BP.  
 XX  
 AC AA219349;  
 XX  
 DT 05-NOV-1999 (first entry)  
 XX  
 DE M. tuberculosis antigen 3' XP25 DNA sequence.  
 XX  
 KM Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KM immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KM immune response; skin test; ss.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN W09942076-A2.  
 PD 26-AUG-1999.  
 XX  
 PF 17-FEB-1999; 99WO-US03268.  
 XX  
 PR 05-MAY-1998; 98US-0072967.  
 PR 18-FEB-1998; 98US-0025197.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1999-527409/44.  
 XX  
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX  
 PS Claim 4; Page 171-172; 299pp; English.

The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to  
 CC AAY39225 are used in the exemplification of the present invention.



XX Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;  
SQ

Query Match 2.6%; Score 48.8; DB 20; Length 985;  
Best Local Similarity 44.9%; Pred. NO. 0.16; Indels 5; Gaps 2;

Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

```

OY 773 ttcccgatgagcagaaagaaacccgagcctgagcccgacacgctcagctg 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 175 ttccgagctgagcagcagcagcagcagcagcagcagcagcagcagcagc 234
OY 833 tgcgagcctgagcagcagcagcagcagcagcagcagcagcagcagcagc 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 235 accgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 294
OY 893 tgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 295 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 354
OY 953 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 ggccttgacgagcagcagcagcagcagcagcagcagcagcagcagcagc 414
OY 1013 tgcacatgagcagcagcagcagcagcagcagcagcagcagcagcagc 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 415 atcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474
OY 1069 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 475 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 533
OY 1129 ctacgacatctgagcagcagcagcagcagcagcagcagcagcagcagc 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 593
OY 1189 gctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 gggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653
OY 1249 cgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 713
OY 1309 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 caccatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 767

```

### RESULT 13

AAZ19137 standard; DNA; 985 BP.

AAZ19137;

05-NOV-1999 (first entry)

M. tuberculosis recombinant antigen DNA encoding 3' Xp25.

Antigen; diagnosis; detection; infection; antibody; immunisation;

vacine; immunity; ss.

Mycobacterium tuberculosis.

MO9942118-A2.

26-AUG-1999.

17-FEB-1999; 99WO-US03265.

05-MAY-1998; 98US-0072596.

18-FEB-1998; 98US-0024753.

PA (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;

XX WPI: 1999-527416/44.

DR New polypeptide comprising antigenic portions of M. tuberculosis

PT Claim 4; Page 217; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding  
CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
CC polypeptides are useful for detecting M. tuberculosis infection in a  
CC biological sample by detecting antibodies which bind with the  
CC polypeptides, and are useful as vaccines for immunizing against  
CC M. tuberculosis infection. The new detection methods are needed as  
CC current vaccination strategies do not provide 100% immunity.

CC Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 2.6%; Score 48.8; DB 20; Length 985;

Best Local Similarity 44.9%; Pred. NO. 0.16; Indels 5; Gaps 2;

Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

```

OY 773 ttcccgatgagcagaaagaaacccgagcctgagcccgacacgctcagctg 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 175 ttccgagctgagcagcagcagcagcagcagcagcagcagcagcagcagc 234
OY 833 tgcgagcctgagcagcagcagcagcagcagcagcagcagcagcagcagc 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 235 accgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 294
OY 893 tgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 295 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 354
OY 953 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 ggccttgacgagcagcagcagcagcagcagcagcagcagcagcagcagc 414
OY 1013 tgcacatgagcagcagcagcagcagcagcagcagcagcagcagcagc 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 415 atcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474
OY 1069 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 475 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 533
OY 1129 ctacgacatctgagcagcagcagcagcagcagcagcagcagcagcagc 1188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 593
OY 1189 gctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 gggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653
OY 1249 cgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 713
OY 1309 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 caccatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 767

```

### RESULT 14

AAZ08699 standard; DNA; 12152 BP.

AAZ08699;

DR	26-SEP-2001	(first entry)
XX	DE	M. carbonacea DNA encoding Everninomicin biosynthetic enzymes.
XX	KW	Everninomicin; antibiotic; bottle-neck gene; orthomycin; fermentation; ds.
XX	OS	Micromonospora carbonacea var. africana.
XX	FT	Key
XX	FT	Location/Qualifiers
XX	FT	complement (189..1064)
XX	FT	/*tag- a
XX	FT	/product- "ORF1"
XX	FT	complement (1069..1073)
XX	FT	/*tag- b
XX	FT	1184..2767
XX	FT	/*tag- c
XX	FT	/product- "ORF2"
XX	FT	complement (2683..3753)
XX	FT	/*tag- d
XX	FT	/product- "ORF3"
XX	FT	complement (3766..4276)
XX	FT	/*tag- e
XX	FT	/product- "ORF4"
XX	FT	complement (4280..4284)
XX	FT	/*tag- f
XX	FT	complement (4526..5368)
XX	FT	/*tag- g
XX	FT	/product- "ORF5"
XX	FT	complement (5392..6147)
XX	FT	/*tag- h
XX	FT	/product- "ORF6"
XX	FT	complement (6152..6156)
XX	FT	/*tag- i
XX	FT	complement (6194..7282)
XX	FT	/*tag- j
XX	FT	/product- "ORF7"
XX	FT	complement (7280..8133)
XX	FT	/*tag- k
XX	FT	/product- "ORF8"
XX	FT	complement (8141..8145)
XX	FT	/*tag- l
XX	FT	complement (8254..9318)
XX	FT	/*tag- m
XX	FT	/product- "ORF9"
XX	FT	complement (9324..9328)
XX	FT	/*tag- n
XX	FT	9575..10504
XX	FT	/*tag- o
XX	FT	/product- "ORF10"
XX	FT	9568..9571
XX	FT	/*tag- p
XX	FT	10584..11585
XX	FT	/*tag- q
XX	FT	/product- "ORF11"
XX	PD	WO200151639-A2.
XX	XX	19-JUL-2001.
XX	XX	12-JAN-2001; 2001WO-US01187.
XX	XX	12-JAN-2000; 2000US-0175751.
XX	PA	(SCHE ) SCHERING CORP.
XX	PI	Hosted by, Horan AC, Wang TX;
XX	PI	WPI; 2001-442147/47.
XX	DR	P-PSDB: AAU04901, AAU04902, AAU04903, AAU04904, AAU04905, AAU04906
XX	DR	AAU04907, AAU04908, AAU04909, AAU04910, AAU04911.
XX	TX	New nucleic acid molecules encoding everninomicin pathway gene

[illegible]

XX 15-JAN-2002 (first entry)  
DT Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
DE Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.  
XX Mycobacterium tuberculosis.  
OS US6294328-B1.  
PN  
XX  
PD 25-SEP-2001.  
XX  
PF 24-JUN-1998; 98US-0103840.  
XX  
PR 24-JUN-1998; 98US-0103840.  
XX  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX  
DR WPI; 2001-647261/74.  
XX  
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
PT determining the nucleotide sequence of the strain at positions in the  
PT genome corresponding to positions where M. tuberculosis strains CDC  
PT 1551 and H37Rv differ  
XX  
PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
XX  
CC The invention relates to evaluating strain variation within and between  
CC different populations of the tuberculosis bacterial pathogen,  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (A199683) and  
CC H37Rv (A199682). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at [seqdata.uspto.gov/sequence.html?DocID=6294328B1](http://seqdata.uspto.gov/sequence.html?DocID=6294328B1).  
XX  
SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 2.6%; Score 48; DB 22; Length 4403765;

Best Local Similarity 45.6%; Pred. No. 1.1; Mismatches 245; Indels 3; Gaps 1;  
Matches 208; Conservative 0;

OY 860 gccggggggcgtaccgggtgatacaacgtagtggcgagacccgagtgacatt 919  
DB 3963782 GTCGGCAACGGGAACCTTGCAACGCAACCTGGGACGCGCAATACCGCACTCCAAC 3963723  
OY 920 gaggtcgactggaacagggcaactacacagggccgacatcagacagaggaac 979  
DB 3963722 GTCGGGTTCCGGGAATACCGCAACAAACGTCGGCACCGGAACCGCGCAAC 3963663  
OY 980 aacgtggccgctatccgagcatctacttcggtgacactggggcgccgtgacagagaat 1039  
DB 3963662 ATCGCGCGCGGGGAACACGCGACGCAACATGGGCTTCGGCAACACGCGCATTTGGCAAC 3963603  
OY 1040 teggagttccgagcggtgagag---agctgtccgagcgtgcagagagctgagcgtc 1096  
DB 3963602 ATCGGTTTCGGCAACACCGGCAACGCGCAACATCGGCTTCGGGCTCACCGGCAACACCAAG 3963543  
OY 1097 acgccgatacagacagggccgctggaatgcccgtacagacatctggtccgtcacg 1156  
DB 3963542 GTGGGCAATCGGTGGGCTGAACTCGGCGAGCGGAACATCGGCTTGTCACTCGGGCAAC 3963483  
OY 1157 aattccgcaacggtacagcgcgccgagctgagtatctgctgaactggaacgagc 1216

DB 3963482 AACAACTCGGTTCTTCAACTCCGGCAACGCAACCTGGGTATCGGCAACTCCAGCGAC 3963423  
OY 1217 ggcgtgagtcgcgggagccgagccggtgaccacgtggaactggccggggccaacttg99aa 1276  
DB 3963422 GCGAATGTGGGCATCGGAACCTCGGGGCGCACGTTGCTGTGGCGGCCATATAAC 3963363  
OY 1277 gctgtgaltgccgactgggactggaattacatcgcc 1312  
DB 3963362 ACCGGTTTCGGGAACCTCGGTTCCGCTCAACACGGGC 3963327

Search completed: August 19, 2002, 13:37:08  
Job time: 13327 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 09:53:16 ; Search time 130.83 Seconds  
(without alignments)  
3524.070 Million cell updates/sec

Title: US-10-003-759-1

Perfect score: 1877  
Sequence: 1 gggcgagtcgaacgctcgggc.....tctccgacacttaaacac 1877

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/prodata/2/ina/5a\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5b\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/5a\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/5b\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131.8	7.0	1116	4 US-09-104-308-2	Sequence 2, Appl
2	131.8	7.0	1116	4 US-09-321-981-2	Sequence 2, Appl
3	105.8	5.6	1716	4 US-09-321-981-4	Sequence 4, Appl
4	49	2.6	756	1 US-08-642-255-50	Sequence 50, Appl
5	48.8	2.6	985	4 US-09-056-556-182	Sequence 182, App
6	48	2.6	4403765	4 US-09-103-840A-2	Sequence 2, Appl
7	47.6	2.5	30001	1 US-08-125-468-1	Sequence 1, Appl
8	47.6	2.5	30001	2 US-08-474-933-1	Sequence 1, Appl
9	47.2	2.5	1908	2 US-08-173-508-1	Sequence 1, Appl
10	47.2	2.5	1908	2 US-08-265-310-1	Sequence 1, Appl
11	47.2	2.5	1508	3 US-08-951-742-1	Sequence 1, Appl
12	46.8	2.5	15872	4 US-09-105-537-1	Sequence 1, Appl
13	46.2	2.5	1288	1 US-08-440-856A-9	Sequence 9, Appl
14	46	2.5	44377	2 US-08-804-227C-7	Sequence 7, Appl
15	46	2.5	44377	2 US-08-804-198-1	Sequence 1, Appl
16	45.2	2.4	1291	3 US-08-997-897-1	Sequence 1, Appl
17	45.2	2.4	1291	4 US-09-156-836B-1	Sequence 1, Appl
18	45.2	2.4	2852	4 US-09-056-556-203	Sequence 203, App
19	44.6	2.4	1176	2 US-08-387-942C-17	Sequence 17, Appl
20	44.6	2.4	11230	4 US-09-105-537-32	Sequence 32, Appl
21	44.6	2.4	12588	2 US-08-387-942C-1	Sequence 1, Appl
22	44.6	2.4	36778	4 US-09-105-537-5	Sequence 5, Appl
23	44.6	2.4	38506	3 US-09-320-878-19	Sequence 19, Appl
24	44.4	2.4	2712	3 US-09-025-691-4	Sequence 4, Appl
25	44	2.3	2064	1 US-08-343-428-1	Sequence 1, Appl
26	43.6	2.3	1017	4 US-08-849-751-1	Sequence 1, Appl
27	43.6	2.3	1017	4 US-09-478-816-1	Sequence 1, Appl

28	43.4	2.3	1500	4 US-09-593-711A-10	Sequence 10, Appl
29	43	2.3	1273	4 US-09-319-892-3	Sequence 3, Appl
30	43	2.3	1333	4 US-09-372-422A-9	Sequence 9, Appl
31	42.8	2.3	599	1 US-08-584-226-3	Sequence 3, Appl
32	42.4	2.3	20235	1 US-07-642-734C-3	Sequence 3, Appl
33	42.4	2.3	20235	3 US-08-439-009A-3	Sequence 3, Appl
34	42.2	2.2	1158	4 US-09-372-422A-21	Sequence 21, Appl
35	42	2.2	71989	4 US-09-443-501A-2	Sequence 2, Appl
36	41.8	2.2	432	1 US-08-642-255-48	Sequence 48, Appl
37	41.6	2.2	1155	2 US-08-387-942C-7	Sequence 7, Appl
38	41.4	2.2	669	3 US-08-913-264-2	Sequence 2, Appl
39	41.4	2.2	2563	1 US-08-608-241-5	Sequence 5, Appl
40	41.4	2.2	2563	2 US-08-922-182-5	Sequence 5, Appl
41	41.4	2.2	2563	2 US-08-919-953-5	Sequence 5, Appl
42	41.4	2.2	18994	1 US-08-459-586-4	Sequence 4, Appl
43	41.4	2.2	18994	2 US-08-282-696-4	Sequence 4, Appl
44	41.4	2.2	4411529	4 US-09-103-840A-1	Sequence 1, Appl
45	41	2.2	1182	4 US-09-385-028-19	Sequence 19, Appl

## ALIGNMENTS

RESULT 1  
US-09-104-308-2  
; Sequence 2, Application US/09104308  
; Patent No. 6187577  
; GENERAL INFORMATION:  
; APPLICANT: Jones, Brian E.  
; APPLICANT: Van der Kleij, Wilhelmus A.H.  
; APPLICANT: Van Solingen, Piet  
; APPLICANT: Weyler, Walter  
; TITLE OF INVENTION: No. 6187577el Cellulase Producing Actinomyces.  
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genencor International, Inc.  
; STREET: 925 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1013  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/104,308  
; FILING DATE: 24-JUN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/974,042  
; FILING DATE: 19-NOV-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stone, Christopher L.  
; REGISTRATION NUMBER: 35,696  
; TELECOMMUNICATION INFORMATION: GC539  
; TELEPHONE: 650-846-7555  
; TELEFAX: 650-845-6504  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1116 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-104-308-2  
Query Match 7.0%; Score 131.8; DB 4; Length 1116;  
Best Local Similarity 50.7%; Pred. No. 1.7e-21;  
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

```
QY 807 ctgagccgagccgaccgctcagctgtgctgagcgtgagacgagcgagatgtgctgagcggg 866
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 CGGCTCCGGCCGCAAGGCAACAGAGATCTGCGACCGCTACGGCACACACACATCCAGG 136
QY 867 ggcgtaccggtgatacaacagatggtgagcgagagaccgagccagtcagtgaagtcg 926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 ACCGGAGACGGTGTGAGAMACACCGCGGGGACACAGCGCCACCATGTGATCAATGTGA 196
QY 927 gactgaaacgggcaactcagatcacagggccgagtcagcaagaagcaaacgctgg 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 CCGGCAACGGTTTCGAGATACCCAGGCCAGCGTTCTGGTCCGACCAACAGGCCGCCGA 256
QY 987 ccgcatatccggcatctactcgtgtgtccactggtgagcggtcgtcagagcaatccggat 1046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 AGTCTATCCCTCGTCTAGAGACGGGCTGCCATCACTACAGCACTGGCGCCCGCACGACGC 316
QY 1047 tgcgcgagcggtgaggaactgttcgagctgtgcagagagcttgaagctcaagcgatca 1106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 TGCCTATGCGGATCAGCTCATGCGCAGCGCCGACAGCATGTCTCTACCGCTACACCG 376
QY 1107 cgaacggcgctggaattgcccctacagacatctgttcagttcccgctcaagcaatccggca 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 GCAACGGCGCTCTACAGCGCGCGCTACGACATCTG-----CTGGACCCGACACCCCGCA 430
QY 1167 acggtctacagcggtgagccgagctgatactgtgtcgtgaactggaacggtgagctgagc 1226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 CCAACGGGGTGAACCGGACCGCATCATGTGATCTGTTCAACCGGGTCCGCGCCGCTCAGC 490
QY 1227 cggagcgagcggtgagcagcggtggaactgtgcggagcgagccactgtggaagctgtgagt 1286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 491 CCACTCGGTTGCGGCTGCGACAGGCCACACTCGCGCGCCCACTGGGAGGTGTGAACGC 550
QY 1287 ccgacttggagctggaattacatcgtcctacccggtgagcaagacccaacagctgcgtgagcg 1346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 551 GCAGCAACGGTTGGAACGAGATGATCTCTCTGCGCGCCCTCCGCGATCAGCAGCTGGA 610
QY 1347 agcttgaacggagggctctatcgaagacggtgtgcggcggtctacatccggtcgagagt 1406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 611 GCTTCGAGGTCAAGAGATGTGTGACCAAGCGCTCAACCCAGCGCTGCGCACCCCGACT 670
QY 1407 ggtatctgcatgtggtggaagcggtctcgaactctgtagggagcggtgagctgtcgaa 1466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 671 GGTACCTCACACATCCAGCGGGCTTCGAGCGGTGAGCGGTGAGGGGCGCACCGGTCTGCCG 730
QY 1467 ggcgcgatttctcgttaacggtgca 1491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 731 TGAACCTGTTCTCTCCGCGGTGAA 755
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
US-09-321-981-2
; Sequence 2, Application US/09321981
; Patent No. 6287839
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: No. 6287839el Cellulase Producing Actinomycetes,
; FILE REFERENCE: GC540-2
; CURRENT APPLICATION NUMBER: US/09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
```

```
; OTHER INFORMATION: Nearest "neighbor" = Streptomyces
; OTHER INFORMATION: thermovibrio
US-09-321-981-2

Query Match      7.0%; Score 131.8; DB 4; Length 1116;
Best local Similarity 50.7%; Pred. No. 1.7e-21;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

QY 807 ctgagccgagccgaccgctcagctgtgctgagcgtgagacgagcgagatgtgctgagcggg 866
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 CGGCTCCGGCCGCAAGGCAACAGAGATCTGCGACCGCTACGGCACACACACATCCAGG 136
QY 867 ggcgtaccggtgatacaacagatggtgagcgagagaccgagccagtcagtgaagtcg 926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 ACCGGAGACGGTGTGAGAMACACCGCGGGGACACAGCGCCACCATGTGATCAATGTGA 196
QY 927 gactgaaacgggcaactcagatcacagggccgagtcagcaagaagcaaacgctgg 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 CCGGCAACGGTTTCGAGATACCCAGGCCAGCGTTCTGGTCCGACCAACAGGCCGCCGA 256
QY 987 ccgcatatccggcatctactcgtgtgtccactggtgagcggtcgtcagagcaatccggat 1046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 AGTCTATCCCTCGTCTAGAGACGGGCTGCCATCACTACAGCACTGGCGCCCGCACGACGC 316
QY 1047 tgcgcgagcggtgaggaactgttcgagctgtgcagagagcttgaagctcaagcgatca 1106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 TGCCTATGCGGATCAGCTCATGCGCAGCGCCGACAGCATGTCTCTACCGCTACACCG 376
QY 1107 cgaacggcgctggaattgcccctacagacatctgttcagttcccgctcaagcaatccggca 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 GCAACGGCGCTCTACAGCGCGCGCTACGACATCTG-----CTGGACCCGACACCCCGCA 430
QY 1167 acggtctacagcggtgagccgagctgatactgtgtcgtgaactggaacggtgagctgagc 1226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 CCAACGGGGTGAACCGGACCGCATCATGTGATCTGTTCAACCGGGTCCGCGCCGCTCAGC 490
QY 1227 cggagcgagcggtgagcagcggtggaactgtgcggagcgagccactgtggaagctgtgagt 1286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 491 CCACTCGGTTGCGGCTGCGACAGGCCACACTCGCGCGCCCACTGGGAGGTGTGAACGC 550
QY 1287 ccgacttggagctggaattacatcgtcctacccggtgagcaagacccaacagctgcgtgagcg 1346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 551 GCAGCAACGGTTGGAACGAGATGATCTCTCTGCGCGCCCTCCGCGATCAGCAGCTGGA 610
QY 1347 agcttgaacggagggctctatcgaagacggtgtgcggcggtctacatccggtcgagagt 1406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 611 GCTTCGAGGTCAAGAGATGTGTGACCAAGCGCTCAACCCAGCGCTGCGCACCCCGACT 670
QY 1407 ggtatctgcatgtggtggaagcggtctcgaactctgtagggagcggtgagctgtcgaa 1466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 671 GGTACCTCACACATCCAGCGGGCTTCGAGCGGTGAGCGGTGAGGGGCGCACCGGTCTGCCG 730
QY 1467 ggcgcgatttctcgttaacggtgca 1491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 731 TGAACCTGTTCTCTCCGCGGTGAA 755
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
US-09-321-981-4
; Sequence 4, Application US/09321981
; Patent No. 6287839
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: No. 6287839el Cellulase Producing Actinomycetes,
; FILE REFERENCE: GC540-2
; CURRENT APPLICATION NUMBER: US/09/321,981
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
```







```

RESULT 7
US-08-125-468-1/c
; Sequence 1, Application US/08125468
; Patent No. 5589385
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Pantlin, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmids
; TITLE OF INVENTION: useful therein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08-125,468
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match 2.5%; Score 47.6; DB 1; Length 30001;
Best Local Similarity 45.3%; Pred. No. 0.056;
Matches 253; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

OY 904 gaccgcccgatcatatgtatgagtctggaactgtaaaacgggcaacttacatcacacgggcga 963
Db 25022 GACCCGCAACGGCTTCGCCCTCGCGCGGCGCGCGGCGGCGGCGATGTTGCTGTGAGAGCGCGGA 24963
OY 964 tcaacacaacggcaacaacacgttgcgcgcataccggccatctactcttggtggtgcacatggg 1023
Db 24962 GCACCGCATCCGGCGCGCGCGCGCGCACATCTACGCCGAGATCACCGGCTACCCCGCGCGCTG 24903
OY 1024 cgcctgcacagacaattcgggaattgcgcgcgcgcgttcgacagagactgtgccgactgacgc 1083
Db 24902 CAACCTCTCTACACATGACCGCGGCTCGCGGCTCGACGGGAAACGAGATGGCCGAGGCTATCCG 24843
OY 1084 gaggctggaacgtccacgcgcatacagacgcgcgcgcgttggaattgcgcctacgaacatctggt 1143
Db 24842 GGTCT---GCGCTGACCGCAGCGCGGATGAACCCCGAGAGAGGTGGCTACTCTCAACGGCGCA 24786
OY 1144 cagtcccgttcacgaattccgcgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1203
Db 24785 CGGCTCGCGGACCA---GCAGAACGACCGCGCACGAGACCGCGCGGCTTCAAGAAAGACCT 24729
OY 1204 gaactcggaacgcgcgcgtgatgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1263
Db 24728 GGGCGACACCGCTTACCGGGGTGCGGCTCAGCTTGATCAATGCATGATCGGACACTTGCT 24669

```

```

OY      1264  gccacccctggggaagcgttgtagcgagcgcgagcgggactgtaattacatgccataccgagcgcaac 1323
Db      24608  gggcgccgcttccgaccttgagcttcgcccccacacggcgctggcgcatgagcacatcggctgctg 24609
OY      1324  gagcgccacacagtcggttagcgagcgttgagccttgaaggccttcacatgcagcagcggtgcg 1383
Db      24608  gccggccacacggcgcaaccttcacagtgcgcgatcccgatccgagctgcgactgcgactgcgcgct 24549
OY      1384  ccgagcgctaccatccgagcgcgagttgatatgcattgcgttgtagagaagcggtctgcgaactctg 1443
Db      24548  ggctgcgcgcgggaacacgccccgctggcgacagctgctcacaacgctggcgagcgcttccgcgctt 24489
OY      1444  ggaaggcgagggcgcgctgtc 1461
Db      24488  ccagagccgcatgctgct 24471

RESULT      8
US-08-474-933-1/C
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strachy, Nancy
; APPLICANT: Pantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlorotetracycline and tetracyline formation and cosmids
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,933
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,468
; FILING DATE: 22-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305
; INFORMATION FOR SEQ ID NO.: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match      2.5%; Score 47.6; DB 2; Length 30001;
Best Local Similarity 45.3%; Pred. No. 0 056;
Matches 253; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

OY      904  gacgcgccagtcgcatgagctgcgactgaaaacgggcaacttcagatcacacgagcgcca 963

```

```

Db 25022 GACCCGCAAGCGCTTCGCCCTGGCGAGGGCGCGGATGTCGTGAGAGGCGGGA 24963
Qy 964 tcacgacaacggaacacagctgacgacatccgacatctcagctgacacgagga 1023
Db 24962 GCACGGGATCCGGCGGGCGGCGACATCTACGGGAGTCCAGCGGTACCGCGGGCTC 24903
Qy 1024 cgcctcagacgaatcggagatcgccgagcgctgacagagctgacagctgacac 1083
Db 24902 CAACTCTCTACAGATGACGGGGCTCGGGGTCCAGCGGAACGAGATGGCGGATCCG 24843
Qy 1084 gacgtgagacgtcacgacgacatccgacgacgacgacgacgacgacgacgac 1143
Db 24842 GGTCTCGACCGGACCGGATGAAACCGGAGAGTCTGCTACGCTCAACCGCA 24786
Qy 1144 cagtcacgacgaatccgacgacgacgacgacgacgacgacgacgacgacgacg 1203
Db 24785 CGGCTCGGGAGCA---GCAGACGACCGGCGGACGAGACCGCGCTTCAAGAGAGCT 24729
Qy 1204 gaactgaaacgagcgctgacgagcgagcgagcgagcgagcgagcgagcgagcg 1263
Db 24728 GGGCGACACCGCTTACCGGGTGGCGGTGCTGATCAAGTGTGATGAGACACTGCG 24669
Qy 1264 ggcacactgagagctgtagctgacgacgagcgagcgagcgagcgagcgagcg 1323
Db 24668 GGGCGGCTGCTGGCGCTGAGTTCGCCCGCACCGCGCTGGCGGATGAGCACTGCTG 24609
Qy 1324 gacgacacacgctgctgacgagcgagcgagcgagcgagcgagcgagcgagcgag 1383
Db 24608 GCGCGCCACGCGCACCTGACGAGTCCCGATCCCGAGTGGAGACTGACACTGCGCGT 24549
Qy 1384 cgcgagctacatccgagcgagctgtagctgacgagcgagcgagcgagcgagcg 1443
Db 24548 GGTGGGGCGGGAACACGCGGTGGCGACGCTGCTCAACGCTGGCGGCTTCCGCGCT 24489
Qy 1444 ggaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1461
Db 24488 CCAGAGCGGATGCTGCT 24471

```

```

RESULT 9
US-08-173-508-1
; Sequence 1, Application US/08173508
; Patent No. 5616485
; GENERAL INFORMATION:
; APPLICANT: Bartfeld, Daniel
; APPLICANT: Butler, Michael J.
; APPLICANT: Hadary, Dany
; APPLICANT: Jenish, David
; APPLICANT: Krieger, Timothy
; TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED
; TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/173,508
; FILING DATE: 23-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/125/CACO
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1908 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 146..1759
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 146..148
; OTHER INFORMATION: /note="Met at position -39
; OTHER INFORMATION: represents fmet"
; NAME/KEY: sig_peptide
; LOCATION: 146..262
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 263..1756
; US-08-173-508-1

Query Match 2.5%; Score 47.2; DB 1; Length 1908;
Best Local Similarity 46.7%; Pred. No. 0.038;
Matches 185; Conservative 0; Mismatches 208; Indels 3; Gaps 1;

Qy 1080 gcaagagcttgagcgtacacgacgacgacgacgacgacgacgacgacgacgacg 1139
Db 159 GCATACGGGCGGAGGCGGACCGGCTTGGGACGCGCGGACGACACTGTGTACCGCACGCTGA 218
Qy 1140 gttcagctccgctcagcaaatccggaacgagctacagcgagcgagcgagcgagcgag 1199
Db 219 TCGCCGCGCGCTCTGCGACCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278
Qy 1200 ggtcgaacttgaaacgagcgagctgacgagcgagcgagcgagcgagcgagcgagcg 1259
Db 279 GCGGAGAGTGGGAGCGGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 338
Qy 1260 cgcgagcgacgtgagagctgtagctgacgagcgagcgagcgagcgagcgagcgag 1316
Db 339 CGGCGATCGACTGGGAGGAGCTGGCGGACCGGACGAGCACTGCGCAAGCCCATCTCACTGCG 398
Qy 1317 ggcgacgagcgacccacacgagctgtagcgagcgagcgagcgagcgagcgagcgag 1376
Db 399 GCTACGTCACGGTCCGATGAGACTACGCCAAGCCGTACGCAAGCATCAGGCTCGCG 458
Qy 1377 cgttcgcccgcggtacatccgagcgagctgtagctgacgagcgagcgagcgagcgag 1436
Db 459 TCGACCGCATGCGCACACCGGAAACGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 518
Qy 1437 aactctgagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1472
Db 519 CCGGCGGTCCCGCGGCTCGCGGCTGCTGCTCCCGG 554

RESULT 10
US-08-265-310-1
; Sequence 1, Application US/08265310
; Patent No. 5856166
; GENERAL INFORMATION:
; APPLICANT: Bartfeld, Daniel
; APPLICANT: Butler, Michael J.
; APPLICANT: Hadary, Dany
; APPLICANT: Jenish, David
; APPLICANT: Krieger, Timothy
; APPLICANT: Malek, Lawrence T.
; APPLICANT: Soostmeyer, Gisela
; APPLICANT: Walczyk, Eva
; APPLICANT: Kryszman, Phyllis

```







CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-198-1

Query Match 2.5%; Score 46; DB 2; Length 44377;  
Best Local Similarity 47.6%; Pred. No. 0.14;  
Matches 136; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
QY 779 gatgggacaaggaagaaagaaacggagccttgagcccgagccgacgctcgagctgtcgga 838  
DB 18968 GACGGCGGCGTCCGCTTGCGCGGAGCACCGGTGCTGATCCGCTCGCGCTGCGGCGGA 18909  
QY 839 cgtggaacgagcgagatgtgagcgagggggcgctaccgggtgatacaacaagatagggc 898  
DB 18908 GAGCAGCGCGTGTGTCAGCACCAGCGGCGGTGTCACACGACGTCAGCGGCTGTCGTC 18849  
QY 899 gcgagagaccgcccagtgcatltgaggtcgagactggaacgggcaacttcaagatcacacg 958  
DB 18848 GGGGATGCTGTCGAGCAGCGCGGCGCAGGCGTCCGCTCCGTGACGTCACAGGCCGTAC 18789  
QY 959 gccgatacagacaagcaacgaacgtagcgccatccgggcaacttctggtgtgcaac 1018  
DB 18788 GCGGACCTGCGCACCCGACTCGGTCAACTCGGCGACCACTCGATCGACCGCGGCTGCTC 18729  
QY 1019 tgggagccttgacgagcaattcggaattgcgagcgagcgatgcag 1064  
DB 18728 GGGACCGCGCTGCCCCACGACGAGATGGCGCACCTTGTGCGG 18683

Search completed: August 19, 2002, 13:28:03  
Job time: 1287 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 09:47:01 ; Search time 3821.57 Seconds

(without alignments)  
6629.156 Million cell updates/sec

Title: US-10-003-759-1

Perfect score: 1877

Sequence: 1 gggcgagtcgaacgtcggc.....tctccgaccttaacaac 1877

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estcin:\*  
4: em\_estmu:\*  
5: em\_estlov:\*  
6: em\_estcpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estcl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.8	3.2	925	12	CNS0091P
2	58.6	3.1	935	12	CNS006XK
3	57.2	3.0	925	12	CNS0091P
4	55.6	3.0	932	12	CNS00720
5	54.6	2.9	932	12	CNS00720
6	54.4	2.9	964	10	BF253875
7	53.8	2.9	964	12	CNS003WG
8	52	2.8	1009	12	CNS010EW
9	51	2.7	844	12	CNS0052P
10	51	2.7	935	12	CNS006XK
11	50.8	2.7	1180	12	AG063851
12	50.6	2.7	700	12	AG060149
13	50.6	2.7	878	10	BE785566
14	50.4	2.7	1022	9	BE040131
15	50	2.7	894	12	AG060214
16	49.8	2.7	449	10	BE517742
17	49.6	2.6	682	9	AU095988

C 18	49.2	2.6	534	12	A2934730
C 19	49.2	2.6	546	12	A2935386
C 20	49.2	2.6	644	9	AU030637
C 21	48.2	2.6	443	9	AV914669
C 22	48	2.6	446	10	BM373545
C 23	48	2.6	559	10	BE405069
C 24	48	2.6	588	10	BE255157
C 25	48	2.6	604	9	AV941745
C 26	48	2.6	606	10	BE255408
C 27	48	2.6	631	9	AL505254
C 28	48	2.6	635	9	AV932837
C 29	48	2.6	637	9	AV933847
C 30	48	2.6	666	10	BF628036
C 31	48	2.6	719	9	AV935472
C 32	48	2.6	830	10	BF254391
C 33	48	2.6	1101	12	CNS0175Y
C 34	47.8	2.5	307	9	AU070991
C 35	47.8	2.5	578	9	AU094712
C 36	47.8	2.5	623	9	AU032635
C 37	47.6	2.5	517	9	AV933848
C 38	47.4	2.5	407	10	BE475922
C 39	47.2	2.5	637	10	BE484578
C 40	47	2.5	657	10	BI717319
C 41	47	2.5	677	9	AL505169
C 42	46.8	2.5	269	10	BE636900
C 43	46.8	2.5	641	9	AU093820
C 44	46.8	2.5	645	12	CNS01213
C 45	46.8	2.5	669	9	AU096005

## ALIGNMENTS

RESULT 1  
CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence T893 end of BAC #  
DEFINITION BAC119016 of RPT-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL053013.1 GI:4934461  
VERSION AL053013  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 925)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPT-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source  
1..925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPT-98"







[illegible]

RESULT 6

LOCUS BF253875 864 bp mRNA linear EST 22-OCT-2001

DEFINITION HVSMEF0002F20f Hordeum vulgare seedling root EST library HVCNDA0007 (Etiolated and unstressed) Hordeum vulgare cDNA clone

ACCESSION BF253875

VERSION HVSMEF0002F20f, mRNA sequence.

KEYWORDS BF253875

SOURCE BF253875.2 GI:13116780 EST.

ORGANISM barley.

REFERENCE Hordeum vulgare

AUTHORS Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Hordeum.

TITLE 1 (bases 1 to 864)

JOURNAL Ming, R., Close, T.J., Kleinhoofs, A., Wise, R., Begum, D., Frisch, D., Yu

COMMENT On Nov 16, 2000 this sequence version replaced gi:11182980.

CONTACT: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Total hg bases = 333

Seq primer: AATTACCCCTCCTAAAGG

High quality sequence stop: 776.

location/Qualifiers

1..864

/organism="Hordeum vulgare"

/cultivar="Morex"

/db\_xref="taxon:4513"

/clone="HVSMEF0002F20f"

/clone\_id="Hordeum vulgare seedling root EST library

HVCNDA0007 (Etiolated and unstressed)"

/tissue\_type="Seedling root"

/lab\_host="TJc121"

/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI; Seeds were surface sterilized then germinated under asexual conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pluescript SK(-) cDNA phagemids. These steps were performed in the Tuttle Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Ming R, Kleinhoofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html>)"

BASE COUNT 172 a 260 c 271 g 161 t

Query Match	2.9%	Score 54.4;	DB 10;	Length 864;
Best Local Similarity	48.1%;	Pred. No.1.9;		
Matches 154;	Conservative	0;	Mismatches 166;	Indels 0;
Gaps	0;			

  

Oy	1091	acgcctcacgcccagatcacagcgccgtcgtgatgctgcacctaacatcgtgttcagtccc	1150
Dd	195	ACCGCGACTGCAGAACCCGCGCGCAGCTCTGACGACGCCCTTGAGATGTGACCAGCGCTG	254
Oy	1151	gtcacgaattccggcaacgcatcacagcgccgagcgtgatgatcttgctgtaacttgg	1210
Dd	255	AGGAAGAAGTCTGGCGCCGAGAGTCCTGCAACAACGAGCGAGCCCTCAAGCTCTGGGAATTCAGG	314
Oy	1211	aacggcgctgcatgtagcgaggcgccgagccgctgtgacccttggaacttgcggggccacc	1270
Dd	315	GACCTGGTAGACGGAGCCTTGAGCCCACTGCGCGAGCTGCGCTGCAAGGCGGGGAAGAGC	374
Oy	1271	tggagaagctcgttatgcgcactggagacgtcgaaattacatgccttacccgcaacgagccgc	1330
Dd	375	TTCGAGTTCTGGAGCACCCACGATGGGAGAAACACGCGACCTGCTCCAACCTGGACCAACAC	434
Oy	1331	accacgtctcgttgaagcagagcttgaaaccttgaaagggccttcatactgacagcagcgctgcgccgagc	1390
Dd	435	GGCTACTCTGGGAGCGGCGCTGGCTTCAAGGCGCCGCCAACAACCTCATCATCTCTCGCC	494
Oy	1391	tacatcgagccggagtgata 1410	
Dd	495	GACGCGCGGAGATCGTGCCTGA 514	

  

RESULT	7
CNS003WG/c	
LOCUS	964 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR09E09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AJ065254
VERSION	AL065254.1 GI:4942606
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 964)
TITLE	Genoscope.
JOURNAL	Direct Submission
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm. Location/Qualifiers 1..964 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BACR09E09" /note="end : TET3"
FEATURES	
SOURCE	
BASE COUNT	224 a 160 c 60 g 105 t 405 others
ORIGIN	

[illegible]

```

RESULT      8
CONS010EM/c
LOCUS
DEFINITION  CNS010EM 1009 bp DNA linear GSS 26-JUL-1999
            Drosophila melanogaster genome survey sequence SP6 end of BAC
            BAC03P19 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098882
VERSION     AL098882.1 GI:5610493
KEYWORDS
SOURCE
ORGANISM    fruit fly.
            Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1009)
            GenomeScope.
            Direct Submission
            Submitted (22-JUL-1999) GenomeScope - Centre National de Sequençage
            BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            Web : www.genoscope.cns.fr)
COMMENT
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelobAC11.
FEATURES
            source
            1..1009
            location/Qualifiers
            /organism="Drosophila melanogaster"
            /plasmid="pBelobAC11"
            /db_xref="taxon:7227"
            /clone_lib="DrosBAC"

```



COMMENT	FEATURES	BASE COUNT	ORIGIN
<p>Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the <i>Drosophila melanogaster</i> genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP <i>Drosophila melanogaster</i> BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of <i>Drosophila</i> DNA provided by the BDGP from the isogenic strain y2; cn bw sp. The same strain used for the BDGP's Pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a>.</p>	<p>source</p>	<p>257 a 170 c 162 g 96 t 250 others</p>	<p>1.935</p>
	<pre> location/Qualifiers     1.935     /organism="Drosophila melanogaster"     /db_xref="taxon:7227"     /clone_1id="RPCI-98"     /clone="BACRI4N09"     /note="end : T7" </pre>		

Query Match	2.7%	Score 51	DB 12	length 935
Best Local Similarity	28.4%	Pred. No. 8.2		
Matches 120	Conservative	95	Mismatches 208	Indels 0
				Gaps 0
QY 1286	gcccgaactggaactggaattacatcgctaccgcgcgcgaacgaccacacgctcgtgtgaac	1345		
Db 510	GSCMCRRSSGSGTGCSTMCMYMSSSVSCCSCSGTCCGCTSCCKMCSCTYGCCKCGCGC	569		
QY 1346	gagctggaacctgaagagccctcatcgaagcagcggtgcgcgcgcgactatccgcgcgag	1405		
Db 570	GCSFSCSSSSCCSBSTYCTCCTBCTKSSSGCSTGSGCTGCCGGGSGCGCGSGCGCC	629		
QY 1406	tgtatctgcattgcgtggtgagacgaggttcgaactctctgagagggcggtctgcga	1465		
Db 630	GSGCGCGCGGCGSSSGSSGSSGSSGSSGSSSGSGSGSGGSGGCGCGCGCGCGSG	689		
QY 1466	agcgccgcatlcttcctgaatcgtgcagtagcgtctgcacaggggcacgcgtatgacga	1525		
Db 690	CGSCGCGSSGCGSGGCGSGSGCGSCCCSCGCGSGSGSGCGSCCSCCGSGCGSGS	749		
QY 1526	gaagcaccccgctcgaggcgactatgcgggcgcgtctgcgtctgtgctacttgaatgac	1585		
Db 750	SSGSGCCSCGCGCGGMSGCGSGSGSGCGCGCGCGCGCGCGCGCGCGCGCGSGC	809		
QY 1586	gtctcggagaggtctaacgcgcggcggtgtgtgttctctctgagagcgctcttttcgcga	1645		
Db 810	GCGCGCGGSSGCGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGSG	869		
QY 1646	cgagaacttctcggaaaagacaagagagagacgtgtgtaacgacaggtctgactcgcac	1705		
Db 870	CGSGSSGCGSSSGSGSGSGCGCGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	929		
QY 1706	cag 1708			
Db 930	CCG 932			

RESULT 11

AG063851/c

LOCUS

DEFINITION Pan troglodytes DNA, clone: PTB-052M02.F, genomic survey sequence.

ACCESSION AG063851

VERSION AG063851.1 GI:16615653

SOURCE GSS: GSS (genome survey sequence).

Pan troglodytes male lymphoblast DNA, clone\_1lb: PTB Chimpanzee Male

BAC library clone: PTB-052M02.F.

ORGANISM Pan troglodytes

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.	1 (sites)	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H., and Sakaki, Y.	BAC end sequences of library PTB	Unpublished	2 (bases 1 to 1180)	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.	Direct Submission
		Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Shohri-Chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpansecgsc.riken.go.jp, URL: http://bgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)					
		Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.					
		PRIMERS					

FEATURES					
	SOURCE	Location/Qualifiers			
		1..1180	/organism="Pan troglodytes"		
			/dd_xref="taxon:9598"		
			/clone="PTB-052M02.F"		
			/sex="male"		
			/cell_type="lymphoblast"		
			/clone_lib="PTB Chimpanzee Male BAC Library"		
	BASE COUNT	182 a	511 c	453 g	28 t 6 others
	ORIGIN				
	Query Match	2.7%; Score 50.8; DB 12; Length 1180;			
	Best Local Similarity	46.5%; Pred. No. 9.2;			
	Matches 198; Conservative	0; Mismatches 227; Indels 1; Gaps 1;			
QY	171	ggcgcacatgtgcgcgtgtcctcatcagtgatgtgtcgtgaccacagatgttggcgagac	230		
Db	712	GGCGCCCCCGGCGATCCTTCGAACCGGGGCGCCCGCGGCTACCGGCGTGCGCGCGGT	653		
QY	211	cacaactactcttcgcgcgcgcgaacacgtgaagtgcgcgcgtgcgtctctgtgcgttcg	290		
Db	652	GCGGCGGCT-GTGCGGGGGGGGGTTGGTGTGTGGCGGGGCTCCTGTGACAGTCCGCC	594		
QY	291	ccaacgactacgctctccgggatgttgtctactccggtcccaactctcagaacgcatfgaac	350		
Db	593	GTGGGGGTCGCGTCGCCCGCGGGCGGGCGGCGTGCCTGCCCGCCCTTCCTTCCTCG	534		
QY	351	gtgaaagtgcgcgcgtgcgtctctatttcgctatgcgcgaaggaggaactgtgtctcgacc	410		
Db	533	CGGGGGGGGCGCGCGCTCCGGGGCTGCTCGGAGGTGGGGTCGCGGCTTCGTGCG	474		
QY	411	cgc	470		
Db	473	GCGCGTGGGGGTGGGGCGGGCGCGCGCTGGCGTCCGGGGGTCTCCGCGTCCGGCGCC	414		
QY	471	gggtacgggttcgaaagcgaaacgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	530		
Db	413	TTCGTGCGGGCGGGGGCGCGCGCGCGGTTCCTTCGTGCGTGGGGCGGCGTGGCGCG	354		
QY	531	cgtgtgcgtatgt	590		
Db	353	CCGGGGGCGCGCGTGCCTTCGCCCGCTCCCGCGCGGGGGCGCGCGCGCGCGCTCG	294		
QY	591	cgcgcctt 596			
Db	293	CCCCTT 288			

	RESULT	12	
ACG60149	LOCUS	700 bp	DNA linear GSS 03-NOV-2001
Pan troglodytes DNA, clone: PTB-047122.R,	DEFINITION		genomic survey sequence.
ACG60149	ACCESSION		
AGC60149.1 GI:16611379	VERSION		
GSS: GSS (genome survey sequence).	KEYWORDS		
Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male	SOURCE		
BAC library clone:PTB-047122.R.			
Pan troglodytes	ORGANISM		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.	REFERENCE		
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,	AUTHORS		
Totochi, Y., Matanabe, H. and Sakaki, Y.			
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical	TITLE		
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);	JOURNAL		
1-7-22 Shohiro-Chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan	REFERENCE		
(E-mail:chimbess@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,	AUTHORS		
Tel.:81-45-503-9111, Fax:81-45-503-9170)			
Clones are derived from the chimpanzee BAC library PTB This BAC end	COMMENT		
was generated during the R&D process and may have higher chance of			
clone tracking errors.			
PRIMERS			
Sequencing: MJ3Rev			
LIBRARY			
Vector : pKS145			
R.Site 1 : SacI			
R.Site 2 : SacI.			
Location/Qualifiers			
1..700			
/organism="Pan troglodytes"			
/db_xref="taxon:9598"			
/clone="PTB-047122.R"			
/sex="male"			
/cell_type="lymphoblast"			
/clone_lib="PTB Chimpanzee Male BAC Library"			
BASE COUNT 30 a 291 c 334 g 15 t 30 others			
ORIGIN			
Query Match 2.7%; Score 50.6; DB 12; Length 700;			
Best Local Similarity 42.7%; Pred. No. 9.4;			
Matches 243; Conservative 0; Mismatches 323; Indels 3; Gaps 1;			
776 ccgcgtgctgcgaacaggaaagaacgcgtgagcccgagccgcgtcgtgagtgtgc 835			
24 CCCGGTNNNGNNGGNNGANNNGAGGCCCGGAGGGGGGGCGGGCGGCCTGTGGCG 83			
836 ggagcgtggagcgcgcgcgatctgtgcccgggggcgcctaaccggatcaacaactatg 895			
Db 84 GGGNNGGNGGGGGGAGCTTTGGGGCCGNGGGGGGGAAGGGGGGGGGGGGGGGCGC 143			
896 ggcgcggagaccgcgccagtcatctgattgagtgcgactgtgaaaacgggcaacttcagatcca 955			
Db 144 GCCCGGGCGGNGGGGGCGCGGGGGGGGGGGGGGGGTTTCGGGGGGGGCGCGCGMANAGG 203			
956 cgaggcgcgttcgcacaacaggcaaacagctggccgcctctccggcatctacttcttggtgc 1015			
Db 204 GGGGCGCAAGACAGACAAACCACCTCGTGCTATCAAGGGCGGGGCCCNMCCGGGNGGC 263			
QY 1016 cactggggcgcgtcgtgagagcaattcggatctgcgcgcgcgcgtgtgagagctgtcgcac 1075			
Db 264 CGGCCCCCGCCCCCGCCC---GGCGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCG 320			
Y 1076 gtgcgcacgagctgtagcgtcaacggcgaatcaagacgggcgcgtgtaattgcgcctacgac 1135			

[illegible]

[illegible]

RESULT 14	
BE040131/c	
LOCUS	BE040131
DEFINITION	1022 bp mRNA linear EST 07-JUN-2000
DESCRIPTION	OD102C08 OD <i>Oryza sativa</i> cDNA 5' similar to gata-binding transcription factors, mRNA sequence.

ACCESSION	BE040131
VERSION	BE040131.1
GI	8335147

KEYWORDS	EST.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa

**REFERENCE**

1 (bases 1 to 1022)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartioideae; Oryzae; Oryza.

**AUTHORS** Bonnett, H. J., Borchert, C., Brazille, S., Brooks, J., Eaton, M., Ferrer, H., Kawasak, S., McCollough, A., Michalowski, C. B., Palacio, C., Scera, G., Wheeler, M. and Zepeda, G. R.

**TITLE** Functional Genomics of Plant Stress Tolerance

**JOURNAL** Unpublished (2000)

**COMMENT** Contact: Michalowski, C. B.

University of Arizona  
Bio Sciences West room 513, Tucson, AZ 85721, USA  
Tel: 520-621-7982  
Fax: 520-621-1697  
Email: cbm@u.arizona.edu  
An open reading frame exists.  
Location/Qualifiers  
1..1022

BASE COUNT ORIGIN	196 a	285 c	292 g	217 t	32 others
/organism="Oryza sativa"					
/strain="Pokkali"					
/db_xref="taxon:4530"					
/clone_lib="OD"					
/tissue_type="roots"					
/dev_stage="1 week"					
/note="1 d 150mM NaCl"					

Query Match	2.7%	Score	50.4	DB	9	Length	1022
Best Local Similarity	55.0%	Pred. No.	11				
Matches	99	Conservative	0	Mismatches	81	Indels	0
						Gaps	0

**OY** 1211 aaagcgcgcgctgaatgcccgggcgacgcgcgtgtgccaccgttgaaacttgcgcggggccacc 1270  
| | | | | | | | | | | | | | | |  
**Db** 423 AGCAGCGGCTCGTCACGTGGAGCAGCCCTCCGGCGGCAGCCGCCGCCGCCGCCGCCAGC 364

Dbb

QY 1271  363	<pre>ttgggaagtcgtgatgcgcgactcgtgacatgaattacatctcgtaccgagcgcaagaagcc   TGCAGCGGCGGGCTCCTGT</pre>	1330
-----------------------	---	------

**Dy** 1331 accacgctcggttgagcgaagtctggacctgaagcctcatgcagcaagcgtgcgccgcgac 1390  
||||| | | | | ||||| | | | | | | | |  
**Db** 303 ACCACCTTCGGTGCGAGTTGAGTGTGGAACACCACAAGATCGGGCTCCGCCCGC 244

AG060214		984 bp	DNA	140000	CSC-03-NOV-2001
NC050214					
LOCUS					
RESULT	15				

LOCUS	AB000214	634 bp	DNM	LINEAR	935 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-047K07.F, genomic survey sequence.				
ACCESSION	AB060214				

```

AG0600214.1 GI:16611446
GSS: GSS (genome survey sequence).
    desc: "Genome survey sequence"
    email: "blake@blake.net"
    phone: "1408.999.4444"
    address: "1408.999.4444"

```

ORGANISM Pan troglodytes  
SOURCE BAC Library clone:PTB-047K07.F.  
clone\_id:PTB chimpanzee male  
pan troglodytes male lymphoblast DNA

Chordata: Eukaryota; Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia; Eutheria; Primates: Catarrhini; Homnidae; Pan.  
 1 (sites)

AUTHORS	TITLE
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, I. D., Iida, I., Totoki, Y., Watanabe, H. and Sakaki, Y.	BAC end sequences of library PTB

unpublished  
JOURNAL  
REFERENCE  
2 (bases 1 to 894)

**AUTHORS**  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, J. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

CONTACT: Chimpbesegsc, riken.go.jp, URL: <http://ngp.gsc.riken.go.jp/>,  
(E-mail: [Chimpbesegsc.riken.go.jp](mailto:Chimpbesegsc.riken.go.jp), URL: <http://ngp.gsc.riken.go.jp/>,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENT  
clones are derived from the chimpanzee BAC library P13 this BAC end  
was generated during the R&D process and may have higher chance of  
being tracked across

### PRIMERS

Vector : pKSI45  
sequencing : zims  
LIBRARY

```

      : pns1
      : Sac1
R.Site 1 : Sac1
P site 2 : Sac1

```

FEATURES	Location/Qualifiers	Source
1	804	

```
source
1.034
/organism="Pan troglodytes"
/dh_vraf="taylor.9508"
```

```

/az_xr1- caon.000
/clone="PTB-047K07.F"
/sex="male"

```

```

/sex= male
/cell_type="lymphoblast"
/clonotype="CD8 Chimera200 Majo Doc Tibrarui"

```

BASE COUNT	ORIGIN	CRIME-IND	FBI CHIMPANZEE	MALE	BAC LIBRARY
83 a	380 c	398 g	7 t	26 others	

Query Match	2.78;	Score 50;	DB 12;	Length 894;
Best Local Similarity	44.09;	Pred NO	13.	

best local similarity 44.00; freq. no. 12;  
Matches 307; Conservative 0; Mismatches 383; Indels 8; Gaps 3;

782 ggcgacaacgaaaggaaccgagccctgagcccgaaaccgaccgcgtcgcgagctgtcgcgacgc 841

Db 143 GGGGCGCAGGGGCGGGCGGGCGCCGACGCCCGCCCGCGCGCGCGCGCGGCGGACCC 202

842 tgggacgcgcgcgatgttggccgggggcgcctaccgggtgatcaacaacgtatgggcgcg 901

Db 203 CCGGGGGGGCGGACAGACGGGCGACAAAGGACAAACAGTGGCTTGGC 262

902 gagaccgcccagtcatttgtagtgcgactygaacgycacttcacatcacacygcc 961

Db 263 GGTGCGCCNCGGCCGGCGCCCCGGCCGCGCGCGCGCGCGCGCCAGGGCGAAGCGGC 322

QY 962 gatcagacaacgcaacaacgtgcccgtatccggccatctacttcggtgccaactg 102

Db 323 CCGGCGCGCCGNNCCCCGGGCGGGGGGCGGCNCGGCCCGGGAGCCCGCNC CCCNCCGGGCGNNC 382





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:46:46 ; Search time 53.28 Seconds

(without alignments)  
544.112 Million cell updates/sec

Title: US-10-003-759-2

Perfect score: 1446  
Sequence: 1 MNVNRALVLSLLFLGCDW.....FLMEGAGLRSADESVTVO 261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: A\_Geneseq\_032802.\*  
2: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1980.DAT.\*  
3: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1981.DAT.\*  
4: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1982.DAT.\*  
5: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1983.DAT.\*  
6: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1984.DAT.\*  
7: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1985.DAT.\*  
8: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1986.DAT.\*  
9: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1987.DAT.\*  
10: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1989.DAT.\*  
11: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1990.DAT.\*  
12: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1991.DAT.\*  
13: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1992.DAT.\*  
14: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1993.DAT.\*  
15: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1995.DAT.\*  
16: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1995.DAT.\*  
17: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1996.DAT.\*  
18: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1997.DAT.\*  
19: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1998.DAT.\*  
20: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA1999.DAT.\*  
21: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA2000.DAT.\*  
22: /SIDSL1/gcgdata/hold-geneeq/geneeqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1318.5	91.2	260	AAV06369	Rhodothermus marin
2	1318.5	91.2	260	AA184882	Emericella deserto
3	1318.5	91.2	260	AA184347	Amino acid sequenc
4	470.5	32.5	105	AAV06342	Rhodothermus marin
5	400.5	27.7	386	AAV67497	Protein sequence o
6	386.5	27.4	371	AAV06367	Streptomyces sp. E
7	386.5	27.4	371	AAV08473	Actinomyces sp. 3
8	396.5	27.4	371	AA184880	Streptomyces livid
9	396.5	27.4	371	AAV84345	Amino acid sequenc
10	396.5	27.4	371	AAV67496	Actinomyces cell
11	343	23.7	381	AAV06368	Streptomyces livid

12	343	23.7	381	21	AA184881	Rhodothermus marin
13	330.5	22.9	429	21	AAV84346	Amino acid sequenc
14	255.5	17.7	261	20	AAV88462	Bacillus lichenifo
15	252.5	17.5	264	20	AAV06370	Erwinia carotovora
16	252.5	17.5	264	21	AA184883	Streptomyces livid
17	252.5	17.5	264	21	AAV84348	Amino acid sequenc
18	224.5	15.5	259	20	AAV06351	Aspergillus aculea
19	224.5	15.5	259	21	AA184864	Aspergillus aculea
20	224.5	15.5	259	21	AAV84329	Amino acid sequenc
21	214	14.8	239	21	AAV06353	Aspergillus kawach
22	214	14.8	239	21	AA184866	Aspergillus kawach
23	214	14.8	239	21	AAV84331	Amino acid sequenc
24	213.5	14.8	254	21	AA184869	Humicola insolens
25	213.5	14.8	254	21	AAV84334	Amino acid sequenc
26	210.5	14.6	246	20	AAV06356	Emericella deserto
27	210.5	14.6	246	21	AA184879	Actinomyces sp. 3
28	210.5	14.6	246	21	AAV84344	Amino acid sequenc
29	210.5	14.6	254	21	AA184868	Humicola grisea EG
30	210.5	14.6	254	21	AAV84333	Amino acid sequenc
31	210.5	14.6	254	22	AAU07558	Humicola grisea en
32	210.5	14.6	254	22	AAU07584	Humicola grisea en
33	210	14.5	254	22	AAU07583	Humicola grisea en
34	203	14.0	255	20	AAV06356	Humicola insolens
35	200.5	13.9	348	20	AAV06362	Gliocladium roseum
36	200.5	13.9	348	21	AA184875	Gliocladium roseum
37	200.5	13.9	348	21	AAV84340	Amino acid sequenc
38	199	13.4	104	20	AAV06357	Streptomyces sp. E
39	194	13.4	253	20	AAV06355	Humicola grisea EG
40	191.5	13.2	234	22	AAU07582	Trichoderma reesei
41	186.5	12.9	244	19	AAV86593	Taraxospora phas
42	186.5	12.9	244	21	AAV44341	T. phaseolina xylo
43	186	12.9	244	20	AAV06359	Fusarium javanicum
44	186	12.9	244	21	AA184872	Fusarium javanicum
45	186	12.9	244	21	AAV84337	Amino acid sequenc

#### ALIGNMENTS

RESULT: 11  
AAV06369 standard; Protein: 260 AA.

06-SEP-1999 (first entry)

Rhodothermus marinus EGIII-like cellulase.

Cellulase: endoglucanase; EGIII: textile; feed additive; baking;

food processing; grain wet milling; pulp; paper.

Rhodothermus marinus.

AAV06369  
124 JUN 20 1999

14-DEC-1998; 98WO-US26552.

16-DEC-1997; 97US-0991720.

(GENV) GENECOR INT INC.

Bower BS, Fowler T, Phillips JT;

WPI; 1999-395187/33.

EGIII like cellulase

Example; Fig 6; 47pp; English.

The present polypeptide represents a full-length sequence of a

CC novel EGIII-like cellulase isolated from genomic DNA using PCR  
CC deduced from a gene sequence isolated from genomic DNA using PCR  
CC primers (see AA59180-91) based on conserved motifs (see AA505325-29)  
CC of *Trichoderma reesei* EGIII cellulase and related enzymes. PCR  
CC has been used to identify novel EGIII-like enzymes, including the  
CC present protein, from bacterial and fungal sources (see AA505331-70).  
CC The sequence shows homology to *T. reesei* EGIII (see AA506330). Also  
CC provided by the invention are vectors, host cells and methods  
CC for the recombinant production of such enzymes, which can be used  
CC in the treatment of cellulose-containing textiles, as feed  
CC additives, in the treatment of wood pulp, in the reduction of  
CC biomass to glucose, in the stone washing of indigo dyed denim, or  
CC as laundry detergent components (all claimed).

Query Match	91.2%	Score 1318.5	DB 20	Length 260
Best Local Similarity	90.3%	Pred. No. 1.7e-118		
Matches 241; Conservative	4;	Mismatches 9;	Indels 13;	Gaps 2

Qy	1	MNVKAVLVLSLLLLFGCDMLPPGDNGKEEPEPEPTVELCGRMDARDVAGRRYINN	60
Db	1	mmvmavlvlsllllffgdcwlfpgdngkepepeptvelcgrmdardvagrlyrrlm	60
Qy	61	VWGAETAOICIEVGLGETGNFTTRDDHDNGNNVAAPAIIEGCHWGA-----CTSNGLP	114
Db	61	vwgaetaoicievglgetgnfttrdhdngnnvaapalyfchaparatarldcaragav	120
Qy	115	RRVQELSVRFSWLTPTPTTRGRMAAADIMPSPTNSGNGSGCAELIMLNMGWYPG	174
Db	115	rrvqelsvrfswltptpttrgrmaaadimpsptnsngsgcaelmlnmgwypmg	173
Qy	175	GSRVAIVELAGATWEWYADDMWYIAYRRTPTTSVSELDLKAIFDADAVARGYIREWY	234
Db	174	gsrvatvelagatwewyadwmwyiayrrtpttsvseldlkaifdadaarylirpwy	233
Qy	235	LHAVETGEIWEKGAGRSADFSYTVQ	261
Db	234	lhavetgeiwegaglrtsadfsylvq	260

RESULT	2
AA14882	2
ID	AA14882 standard; Protein; 260 AA.
XX	
AC	AA14882;
XX	
DT	21-NOV-2000 (first entry)
XX	
DE	Emericella desertoru EGII-1like cellulase.
XX	
KX	Emericella desertoru; Trichoderma reesei; endoglucanase III; EGIII;
KW	cellulase; mutant; enzyme stability; textile treatment;
KW	wood pulp treatment; feed additive; detergent.
XX	
XX	Emericella desertoru.
XX	
PN	WO200037614-A2.
PD	29-JUN-2000.
XX	
PF	12-NOV-1999; 99MO-US26704.
XX	
PR	18-DEC-1998; 98US-0216295.
XX	
PA	(GEMV ) GENENCOR INT INC.
XX	
PI	Mitchinson C, Wendt DJ;
XX	
DR	WPI; 2000-482483/42.
XX	
DT	Novel endoglucanase III or endoglucanase III-like cellulase useful for

treating textiles and wood pulp comprises a substitution or deletion at specified positions in the wild form of endoglucanase III -

CC The prehn sequence is a cellulase related to endoglucanase III (EGIII)  
CC from *Trichoderma reesei*. EGIII-like genes were isolated from genomic DNA  
CC libraries constructed from various microorganisms by PCR. The isolated  
CC genes showed significant homology to EGIII from *T. reesei*. Certain  
CC substitution and deletion mutations have been incorporated into EGIII and  
CC EGIII-like cellulases to produce variant enzymes with improved stability  
CC e.g., increased resistance to temperature stress. The mutants may be used  
CC in textile and wood pulp treatment, as a feed additive, and for reducing  
CC biomass to glucose. They are also useful for stonewashing or indigo dyed  
CC denim and as an agent in laundry and dish detergents.

**SQ Sequence 260 AA;**

Query Match	91.2%	Score	1318.5	DB 21	Length	260
Best Local Similarity	90.3%	Pred. NO.	1.7e-118			
Matches 241	Conservative	4	Mismatches	9	Indels	13
					Gaps	2

Qy	1	MNRAVAVYLSLLILFGDMLPEPGDNGKBEPEPEPTVELCGRMDARDVAAGCYRINN	60
Dd	1	mmvmaavvlsllllfgdcmflpdgdngrkepepepepvelcgrmdardvaagryavmm	60
Qy	61	WGAETACIEVGELETGNETITTRADHDGNNVAAYPAIFYGCHWGA-----CTSNGLP	114
Dd	61	vwgetaqclevglgetgnfltrtdhngnmvaaypalyfsgchwaparatarcaragav	120
Qy	115	RRVGLSIVKRSWMLPTTTRGNMAAADIMSPVTSNGNGYSSGAEIMTLNNNGVMPG	174
Dd	121	rtaheld-----vprltcgrwnaadvlwsptcmngysggaelmldmngwmpg	173
Qy	175	GSRYATVELAGATWEVMYADDMNYIAAYRRTPTTSVELDKAFIDDAVARGYIRPEWY	234
Dd	174	gsrvatvelagatwevmvjadcdmnyiaayrrtpttsveldkelfidavarygylrpewy	233
Qy	235	LHAYETGFELMEGAGLRSAFSTVYQ	261
Dd	234	lhavetgfelwegaglrtdadfstvlyq	260

RESULT	3
ID	AAV84347- standard; Protein; 260 AA.
XX	
AC	AAV84347;
XX	
DT	12-JUL-2000 (first entry)
XX	
DE	Amino acid sequence of an endoglucanase III (EgIII)-like cellulase.
XX	
KW	Endoglucanase III; EgIII; EgIII-like cellulase; surfactant stability;
KW	cellulase; textile processing; textile cleaning; stonewashing;
KW	indigo dyed denim; cellulose containing fabric; fabric smoothness;
KW	pill removal; fibrill removal; cotton; cellulosic fibre; dying; detergent;
KW	animal feed; wood pulp; paper; grain; biomass reduction; glucose.
XX	
OS	Rhodothermus marinus.
XX	
PN	WO200014208-A1.
XX	
PD	16-MAR-2000.
XX	
PF	24-AUG-1999; 99WO-US19154.
XX	
PR	03-SEP-1998; 98US-0146729.
XX	
PA	(GENV ) GENENCOR INT INC.
XX	
PI	Fowler T;











PI Fowler T;  
 XX  
 DR WPI: 2000-271052/23.  
 XX  
 PT Novel variant endoglucanase III-like cellulases with improved  
 PT surfactant stability and resistance to temperature stress, useful for  
 PT textile processing or cleaning, treating wood pulp, food and grain, and  
 PT reducing biomass to glucose  
 XX  
 PS Disclosure; Page 65-66; 73pp; English.  
 XX  
 CC The present sequence represents an endoglucanase III (EGIII)-like  
 CC cellulase. The cellulase has homology to the Trichoderma reesei EGIII  
 CC protein. The variant cellulases have improved temperature stability,  
 CC and improved surfactant stability. The variant cellulases and  
 CC compositions containing them are used in textile processing or cleaning,  
 CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel  
 CC or appearance of cellulose containing fabrics (e.g. improving fabric  
 CC smoothness or removing pills and fibrils). The compositions may also be  
 CC used for the removal of immature or dead cotton from cellulosic fibres  
 CC or fabric, which can cause uneven dying. The cellulase may also be used  
 CC in a detergent composition for washing laundry and dishes and in the  
 CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.  
 CC The enzymes may also be used in the reduction of biomass to glucose.  
 XX  
 XX Sequence 429 AA;  
 SQ  
 Query Match 22.9%; Score 330.5; DB 21; Length 429;  
 Best Local Similarity 32.9%; Pred. No. 2.7e-23; Indels 17; Gaps 5;  
 Matches 79; Conservative 34; Mismatches 110;  
 QY 24 DGDNGKPEPEPEPTVELCGRMARDYAGGRYRVINNVWGAETACIEVGLGNTFTTR 83  
 Db 85 dgsaprtngapksyps-----vfngchytqnmrtgslapqcv-tatdtg-ffvtq 131  
 QY 84 ADHNGNNVA--AYPAITFGCHMGACTSNGSLPRVOELSDVRTSWLTPTTGRMNAAY 141  
 Db 132 adgsaprtngapksypsfnvgchytncspgtlprldtvsapsisgytfdgagynasy 191  
 QY 142 DINSPTVNSGNGYSGAELMTIMNMNGVMPGSGSRVATVELAGATVEWYADMDMNYIA 201  
 Db 192 diwtdprardcg--vnteltniwnfrvgpigrpispvgtasvgrtvevsggsgndvl 249  
 QY 202 YRRTPTTSVSELDLKAFFIDDAVARGYIRPEWYLHAVETGFEIMEGGAGLSADFSVTYQ 261  
 Db 250 sfvapsalsgswfdwmfdvratvargiaendwyltsvqagfepwngaglavnsfsstve 309  
 RESULT 14  
 AAW88462  
 ID AAW88462 standard; Protein; 261 AA.  
 XX  
 AC AAW88462;  
 XX  
 DT 10-MAY-1999 (first entry)  
 DE Bacillus licheniformis xyloglucanase.  
 XX  
 KM xyloglucanase; detergent.  
 XX  
 OS Bacillus licheniformis.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 30..261  
 FT /note="mature protein, this region is specifically  
 FT claimed in Claim 19"  
 XX  
 PN WO9902663-A1.  
 PD 21-JAN-1999.  
 PF 01-JUL-1998; 98WO-DK00290.

XX  
 PR 24-OCT-1997; 97DK-0001213.  
 PR 07-JUL-1997; 97DK-0000822.  
 XX  
 PA (NOVO ) NOVO-NORDISK AS.  
 XX  
 PI Bjornvad ME, Jorgensen PL, Outtrup H, Schuelein M;  
 DR WPI: 1999-120866/10.  
 DR N-PSDB; AAX06949.  
 XX  
 CC New enzyme preparation comprising a xyloglucanase with an activity  
 PT of 50 % at pH 7 - useful for improving the properties of cellulosic  
 PT fibres, yarn, (non)woven fabric, and ratting hemp, jute, flax and  
 PT linen fibres  
 XX  
 PS Claim 29; Page 71-72; 87pp; English.  
 XX  
 CC This polypeptide is an alkaline xyloglucanase obtained from  
 CC Bacillus licheniformis ATCC 14580. The enzyme shows optimal  
 CC activity at 60 deg C, and retains 50% of its activity after 20 min  
 CC at 70 deg. The optimum pH for activity is 5.54, with 32% relative  
 CC activity retained at pH 4.59 and 45% at pH 6.49. The coat is  
 CC 16.5/sec on xyloglucan at pH 7.5, Km 1.1 g/l. The ratio of maximum  
 CC xyloglucanase activity to maximum activity on CM-cellulose is at  
 CC least 5:1. An isolated polynucleotide (see AAX06949) encoding the  
 CC xyloglucanase can be utilised in the production of recombinant  
 CC enzymes. Xyloglucanase preparations are useful for improving the  
 CC properties of cellulosic fibres, yarn, (non)woven fabric, and for  
 CC ratting hemp, jute, flax and linen fibres (claimed). They can also  
 CC be used in a process for machine treatment of fabrics, and in  
 CC detergent compositions (claimed). The enzyme preparations exhibit  
 CC high xyloglucanase activity at alkaline pH without essentially  
 CC attacking cellulose or cellulose derivatives.  
 XX  
 XX Sequence 261 AA;  
 SQ  
 Query Match 17.7%; Score 255.5; DB 20; Length 261;  
 Best Local Similarity 30.6%; Pred. No. 2.2e-16;  
 Matches 70; Conservative 41; Mismatches 85; Indels 33; Gaps 9;  
 QY 54 RYRINNVWGAET-----OCIEVGLGNTFTTADPD-----NGNNVAAYPAITF 100  
 Db 45 kyllfnnvwadqdygswqtl-----yhnsdsdmvwvwnpsntstlvxapsivs 94  
 QY 101 GCHW-GACTSNGSLPRVOELSDVRTSWLTPTTGRMNAAYDIFSPVNSGNGYSGA 159  
 Db 95 gwhtegtatgsgfprtlisdqknltkvsysisngcynaaydlwlnhtkawsdsapt 154  
 QY 160 ELMITLMWNGGVMGSGSRVATVELAGATWEVW--YAD---WDMNYAYRRTPTTSVSE 213  
 Db 155 eimtiwln-nlnagpagsvveltsigshwkykylidagggkgnvwnfsflrantgs-an 212  
 QY 214 LDLKAFIDD-AVARGYIRPEWYLHAVETGFEIMEGGAGLSADFSVTYQ 261  
 Db 213 lnrdftnyladsqkysktyssvefgevtgglqnlisnwdcvr 261  
 RESULT 15  
 AAY06370  
 ID AAY06370 standard; Protein; 264 AA.  
 XX  
 AC AAY06370;  
 XX  
 DT 06-SEP-1999 (first entry)  
 DE Erwinia carotovora EGIII-like cellulase.  
 XX  
 KM Cellulase; endoglucanase; EGIII; textile; feed additive; baking;  
 KM food processing; grain wet milling; pulp; paper.  
 XX  
 OS Erwinia carotovora.







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: August 15, 2002, 10:49:16 : Search time 23.32 Seconds

(without alignments)  
273.374 Million cell updates/sec

Title: US-10-003-759-2

Perfect score: 1446  
Sequence: 1 MNVRAVLVLSLLIFGCDW.....FELMEGAGLRSADESVTVQ 261Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1446.5	91.2	260	US-09-216-295-23	Sequence 23, Appl
2	1400.5	27.7	386	US-09-321-981-5	Sequence 5, Appl
3	396.5	27.4	312	US-09-216-295-21	Sequence 21, Appl
4	396.5	27.4	371	US-09-104-308-1	Sequence 1, Appl
5	396.5	27.4	371	US-09-321-981-1	Sequence 1, Appl
6	343	23.7	381	US-09-216-295-22	Sequence 22, Appl
7	255.5	17.7	261	US-09-110-959A-2	Sequence 2, Appl
8	254	17.6	233	US-08-032-848C-12	Sequence 12, Appl
9	246	17.0	263	US-09-216-295-24	Sequence 24, Appl
10	224.5	15.5	259	US-09-216-295-5	Sequence 5, Appl
11	214	14.8	239	US-09-216-295-7	Sequence 7, Appl
12	213.5	14.8	253	US-09-216-295-10	Sequence 10, Appl
13	213	14.7	221	US-08-032-848C-13	Sequence 13, Appl
14	210.5	14.6	246	US-09-216-295-10	Sequence 10, Appl
15	210.5	14.6	253	US-09-216-295-9	Sequence 9, Appl
16	200.5	13.9	348	US-09-216-295-16	Sequence 16, Appl
17	186	12.9	243	US-09-216-295-13	Sequence 13, Appl
18	179.5	12.4	232	US-09-146-770-1	Sequence 1, Appl
19	176.5	12.2	218	US-08-032-848C-10	Sequence 10, Appl
20	176.5	12.2	218	US-08-438-870-10	Sequence 10, Appl
21	176.5	12.2	218	US-08-169-948B-34	Sequence 34, Appl
22	176.5	12.2	218	US-08-448-873-34	Sequence 34, Appl
23	176.5	12.2	218	US-08-382-452D-34	Sequence 34, Appl
24	176.5	12.2	218	US-09-216-295-1	Sequence 1, Appl
25	176.5	12.2	234	US-08-032-848C-9	Sequence 9, Appl
26	176.5	12.2	234	US-08-438-870-9	Sequence 9, Appl
27	176.5	12.2	234	US-09-146-770-3	Sequence 3, Appl

28	176.5	12.2	234	4	US-09-216-295-3	Sequence 3, Appl
29	172	11.9	239	4	US-09-216-295-15	Sequence 15, Appl
30	167	11.5	237	4	US-09-216-295-19	Sequence 19, Appl
31	166.5	11.5	234	4	US-09-146-770-4	Sequence 4, Appl
32	166.5	11.5	234	4	US-09-216-295-4	Sequence 4, Appl
33	166.5	11.5	239	3	US-09-215-042-1	Sequence 1, Appl
34	166	11.3	250	4	US-09-216-295-14	Sequence 14, Appl
35	163.5	11.3	239	4	US-09-216-295-6	Sequence 6, Appl
36	163	11.3	247	4	US-09-216-295-8	Sequence 8, Appl
37	162.5	11.2	223	3	US-08-913-264-1	Sequence 13, Appl
38	162.5	11.2	239	4	US-08-849-751-2	Sequence 2, Appl
39	161.5	11.2	239	4	US-09-478-816-2	Sequence 2, Appl
40	161.5	11.2	239	4	US-09-216-295-12	Sequence 12, Appl
41	159.5	11.0	194	4	US-09-216-295-11	Sequence 11, Appl
42	145	10.0	194	4	US-08-995-280C-2	Sequence 2, Appl
43	141.5	9.8	288	3	US-09-215-042-2	Sequence 2, Appl
44	141.5	9.8	288	3	US-08-591-685-13	Sequence 13, Appl
45	106.5	7.4	368	3		

## ALIGNMENTS

```
RESULT 1
US-09-216-295-23
Sequence 23, Application US/09216295
Patent No. 6287839
GENERAL INFORMATION:
APPLICANT: Mendt, Dan J.
TITLE OR INVENTION: No. 6286328e1 Variant EGIII-Like Cellulase Compositions
FILE REFERENCE: GC555
CURRENT APPLICATION NUMBER: US/09/216,295
CURRENT PUBLISHING DATE: 1998-12-31
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 23
LENGTH: 260
TYPE: PRT
ORGANISM: Rhodothermus marinus
US-09-216-295-23

Query Match
Best Local Similarity 90.3% Score 1318.5; DB 4; Length 260;
Matches 241; Conservative 4; Mismatches 9; Indels 13; Gaps 2;

QY 1 MNVRAVLVLSLLIFGCDWLFPPGDNKKEPEPEPEVELCGRMDARDVAGRYRYINN 60
DB 1 MNVRAVLVLSLLIFGCDWLFPPGDNKKEPEPEPEVELCGRMDARDVAGRYRYINN 60
QY 61 VMGAETACQIEVGLTGNFTITRADHDGNNVAAPAIYFCGHWGA-----CTNSGLP 114
DB 61 VMGAETACQIEVGLTGNFTITRADHDGNNVAAPAIYFCGHWGA-----CTNSGLP 114
QY 115 RRVVELSDVRKSMFLPTITTRGRMAAADIVRSPYTNNGNGSGGAELMIWLNMGVWPG 174
DB 115 RRVVELSDVRKSMFLPTITTRGRMAAADIVRSPYTNNGNGSGGAELMIWLNMGVWPG 174
QY 121 RRAHELD-----VPTITTRGRMAAADIVRSPYTNNGNGSGGAELMIWLNMGVWPG 173
DB 121 RRAHELD-----VPTITTRGRMAAADIVRSPYTNNGNGSGGAELMIWLNMGVWPG 173
QY 175 GSRATVETELAGATWVYADMDMYIAVRRTPPTVSSELDLKAFFIDDAVARGYIREMY 234
DB 175 GSRATVETELAGATWVYADMDMYIAVRRTPPTVSSELDLKAFFIDDAVARGYIREMY 234
QY 174 GSRATVETELAGATWVYADMDMYIAVRRTPPTVSSELDLKAFFIDDAVARGYIREMY 233
DB 174 GSRATVETELAGATWVYADMDMYIAVRRTPPTVSSELDLKAFFIDDAVARGYIREMY 233
QY 235 LHAVETGFEIWEAGAGLRSADESVTVQ 261
DB 235 LHAVETGFEIWEAGAGLRSADESVTVQ 260
QY 234 LHAVETGFEIWEAGAGLRSADESVTVQ 260
DB 234 LHAVETGFEIWEAGAGLRSADESVTVQ 260

RESULT 2
US-09-321-981-5
Sequence 5, Application US/09321981
Patent No. 6287839
GENERAL INFORMATION:
```

```

; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: No. 6268328 Cellulase Producing Actinomycetes,
; FILE REFERENCE: GC540-2
; CURRENT APPLICATION NUMBER: US/09/321,981
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCT/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 386
; TYPE: PRF
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Nearest "neighbor" = Streptomyces
; OTHER INFORMATION: thermovabaceus
US-09-321-981-5
```

```

Query Match      27.7%; Score 400.5; DB 4; Length 386;
Best Local Similarity 39.0%; Pred. No. 5,8e-32;
Matches 90; Conservative 28; Mismatches 104; Indels 9; Gaps 6;
```

```

OY 33 PEPPVELCGRMDARVAGRYVINNVGAEETACIEVLEGTGN-FTTRADHDGNN 91
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 42 PPAQANOICDRGTTTIO-DRYVONNRKGTSTATOCINV---TGNGFETIADGSPVPT 97
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 92 VA--AVALPGCHMGACTNSGLPRRVOELSDVRTSWTLPTTGRMNAAYDIWFSPTV 149
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 98 GAPKSYSPVDCGYHCAPRTLLPMRISSIGSAPSSVSRYTNGVYNAAYDIWLDLP-T 156
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 150 NSGNGYSGGELMTLWLNMGVMPGSGRVATVELAGATWEVYADMDWNTIARRTTPTT 209
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 157 PRNGVAV-REIMIMFNRVGPVOPISPGVTAHVGSRSEWVGTSNGSNVISFLAPSAI 215
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 210 SVSELDKATIDAVAGRYIRPEVYLAHVETGFEWEGAGLSADFSVTV 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 216 SSMSFDVAKVDQAVSHGLATPDWYLTISIAGFEPMEGCTGLAVNSFSNAV 266
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

RESULT 3
; Sequence 21, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328 Variant Egitr-like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 312
; TYPE: PRF
; ORGANISM: Actinomycete 11AG8
US-09-216-295-21
```

```

Query Match      27.4%; Score 396.5; DB 4; Length 312;
Best Local Similarity 39.7%; Pred. No. 1.1e-31;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;
```

```

OY 40 ELGCRMDARVAGRYVINNVGAEETACIEVLEGTGN-FTTRADHDGNNVA--AYP 96
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 34 QICDRGTTTIO-DRYVONNRKGTSTATOCINV---TGNGFETIADGSPVPTGAPKSP 89
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

OY 97 AIYEGCHMGACTNSGLPRRVOELSDVRTSWTLPTTGRMNAAYDIWSPVTSNGNGS 156
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 90 SYVDGCHYGNCAPRTLLPMRISSIGSAPSSVSRYTNGVYNAAYDIWLDLP-TPRTNGVN 148
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 157 GGAELMTLWLNMGVMPGSGRVATVELAGATWEVYADMDWNTIARRTTPTTSVELD 216
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 149 -REIMIMFNRVGPVOPISPGVTAHVGSRSEWVGTSNGSNVISFLAPSAISSFDV 207
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 217 KATIDAVAGRYIRPEVYLAHVETGFEWEGAGLSADFSVTV 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 208 KDEVDQAVSHGLATPDWYLTISIAGFEPMEGCTGLAVNSFSNAV 251
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

RESULT 4
US-09-104-308-1
; Sequence 1, Application US/09104308
; Patent No. 6187577
; GENERAL INFORMATION:
; APPLICANT: Jones, Brian E.
; APPLICANT: Van Der Kleij, Wilhelmus A.H.
; APPLICANT: Van Solingen, Piet
; APPLICANT: Meyler, Walter
; TITLE OF INVENTION: Cellulase Producing Actinomycetes,
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc.
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,308
; FILING DATE: 24-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/974,042
; FILING DATE: 19-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stone, Christopher L.
; REGISTRATION NUMBER: 35,696
; REFERENCE/DOCKET NUMBER: GC539
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7555
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-104-308-1
```

```

Query Match      27.4%; Score 396.5; DB 4; Length 371;
Best Local Similarity 39.7%; Pred. No. 1.4e-31;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;
```

```

OY 40 ELGCRMDARVAGRYVINNVGAEETACIEVLEGTGN-FTTRADHDGNNVA--AYP 96
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 34 QICDRGTTTIO-DRYVONNRKGTSTATOCINV---TGNGFETIADGSPVPTGAPKSP 89
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 97 AIYEGCHMGACTNSGLPRRVOELSDVRTSWTLPTTGRMNAAYDIWSPVTSNGNGS 156
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 90 SYVDGCHYGNCAPRTLLPMRISSIGSAPSSVSRYTNGVYNAAYDIWLDLP-TPRTNGVN 148
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 157 GGAELMTLWLNMGVMPGSGRVATVELAGATWEVYADMDWNTIARRTTPTTSVELD 216
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```



```

DB      213  INIRFTNVLADSKQMLSKTKYVSSVERGETVEFGGTGGINISNMDDVTNR 261

RESULT      8
US-08-032-848C-12
; Sequence 12, Application US/08032848C
; Patent No. 5475101
; GENERAL INFORMATION:
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen A.
; APPLICANT: Weiss, Geoffrey L.
; APPLICANT: Larenas, Edward
; APPLICANT: Lorch, Jeffrey D.
; TITLE OF INVENTION: Purification and Molecular Cloning of
; TITLE OF INVENTION: EG III Cellulase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/032,848C
; FILING DATE: MAR 17 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 742-7356
; TELEFAX: 415 742-7217
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 233 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-032-848C-12

Query Match      17.6%; Score 254; DB 1: Length 233;
Best Local Similarity 32.0%; Pred No. 1.2e-17;
Matches 73; Conservative 39; Mismatches 86; Indels 30; Gaps 10.

QY      54  RYRVINNWGAETAQ-----CIEVGLTGNFTTRADHDGNNVAAYPAYFGCH 103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      16  KYLLNNWVGKDEIKGMQOTIFYNPSISMG---MWMHMPSSPH---SVKAYPSLVSGH 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      104 WCA-CTSNAGCLRQRQELSDVRTSMTLPIITGRRNNAAYDIWFSEYTLNSGNGTSGAGELM 162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      69  WTAGTENSGLCIQJSSNKSITSNVTYSIKATGTNAAYDIWFTTDXKAMWDSPTDELH 128
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      163 IWLN-WNGGVMPGSGSRVATVELAGATWEY---WYADWD---WNYIAYRRTPTTSVSEL 214
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      129 IWLNDTNGG--PAGDYIEIVFLGDSSWMNVFKWMINNADGCGMNVFSPVHTSGTNSAS-L 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      215 DKAATIDDAV-ARGYIRPEWYLHAAVETGELMEGAGLRADFSVTVQ 261
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      186 NIRHFTDVLVOTKQWMSDEKYYISVEFGTEIFGGGQIDITIEWRVDVK 233

```

```

: GENERAL INFORMATION:
: APPLICANT: Mitchinson, Colin
: TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
: FILE REFERENCE: GC555
: CURRENT APPLICATION NUMBER: US/09/216,295
: CURRENT FILING DATE: 1998-12-18
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 24
: LENGTH: 263
: TYPE: PRT
: ORGANISM: Erwinia carotovora
US-09-216-295-24

Query Match 17.0%; Score 246; DB 4; Length 263;
Best Local Similarity 31.6%; Pred. No. 8,6e-17;
Matches 72; Conservative 40; Mismatches 84; Indels 32; Gaps 11;

QY 54 RYRVINWGAETAA-----CIEVGETGFTTTTRADHDGNNVAAYPAIFGCH 103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 48 KYLENNWNGKDEIGMOOTIFNPSISMG---MMWHPSSYH---SVKAYPSLVSGWH 100
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 WCA-CTNSGGLPRRQVELSDVRTSWTLPIITGRNNAAYDIWFSPVTSNGSGAGELM 162
      | | | | | | | : : : : : : : : : : : : : : : : : : : : :
DB 101 WTAGTETNSGLEITQLSSKSTISNVTYSIKATGITMAAYDIWFHTTDXKANDSSFTDEL 160
      | | | | | | | : : : : : : : : : : : : : : : : : : : : :
QY 163 IWLN-WNGVMPGSGSRVATVELAGATWEWYADW-----DWNIAAYRRTPPTTSSEL 214
      | | | | | | | : : : : : : : : : : : : : : : : : : : : :
DB 161 IWLNDNTNG--PAGDIETVFLGDSSMWV-FKGMINADNGGWN-VSPVHTSGTNSAS-L 215
      | | | | | | | : : : : : : : : : : : : : : : : : : : : :
QY 215 DKAATIDAV-AKGIREDWYILHAVETGPELMEGAGLRADFSTVQ 261
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 216 NIRHPTDYLVOTKQWMSDEKYSIVSEFGETEIFGQGDIDITEMRDVK 263
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-09-216-295-5
: Sequence 5, Application US/09216295
: Patent No. 6268328
: GENERAL INFORMATION:
: APPLICANT: Mitchinson, Colin
: TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
: FILE REFERENCE: GC555
: CURRENT APPLICATION NUMBER: US/09/216,295
: CURRENT FILING DATE: 1998-12-18
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5
: LENGTH: 259
: TYPE: PRT
: ORGANISM: Aspergillus aculeatus
US-09-216-295-5

Query Match 15.5%; Score 224.5; DB 4; Length 259;
Best Local Similarity 29.6%; Pred. No. 1.2e-14;
Matches 76; Conservative 36; Mismatches 92; Indels 53; Gaps 12;

QY 40 ELICGWDARDVAGRGYRVINWGAET--AQCIEV--GLETGNFTITRADHDG-NNVA 93
      : : : : : | | : : : | | : : : : : : : : : : : : : : : :
DB 20 QLCDDY--ATYTGAYYITINNINMGKADAGSGCCTTVNSASSAGTSWSTKMMWSGSENSVK 77
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 AYPALIFGCHWGACTSNSGL---PRRQVELSDVRTSWTLPIITG--RNNAAVDIWFSPVT 149
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 78 SY-----ANSGLTFNKKLIVQISQIPTTARMSYDNGIGIRADVAYDLETTADI 124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 NSGNYSGGAGELMTLNNNGVMPGSGSRVATVELAGATWEWY-ADMMWNNTIAYRRTPPT 208
      | | | | | | | | | | | | | : : : : : : : : : : : : : : : :
DB 125 NHHVT-WSGDYELMTLWARYGVOPIGSQIATPATVDGOTWELWANGSGOKTYSFVAPTP 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

OY 209 TSVSELDKAFID-----DAVARGYR-----DEWYLHAVENTGE 243
DB 164 TSV-QGDVNDFFKXYLTQNHGFPASSQYLITQFGTEPPTGCPATLSVSNMSASVQOAGFE 242
OY 244 LMEGGAGLRSADESVTV 260
DB 243 PWONGAGLAVNSFSSTV 259

RESULT 11
US-09-216-295-7
; Sequence 7, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Aspergillus kawachii (2)
US-09-216-295-7

Query Match 14.8%, Score 214; DB 4; Length 239;
Best Local Similarity 32.1%, Pred. No. 1,2e-13;
Matches 78; Conservative 37; Mismatches 84; Indels 44; Gaps 15;

OY 40 ELGCGMDARDVAGGXYRINNWMGAET--AQCIETV--GLEGNFTTIRADHDNG-NNVA 93
DB 20 QLCDQY--ATYTGYYTINNMLMGKDGSGSCCTVNSASSAGTSMSTKMNWNGGENSVK 77
OY 94 AYPAYIFGCHMGACTSNMGL--PRVOELSDV---RTSWTLRPTITGRMNAADWFS 146
DB 78 SY-----ANGSLFNKKLVLSQISIIIPRAKMSYDNTCIRGR--AYDLFTA 121
OY 147 PVTNSGNGYSGCAELIMLNMNGVMPGSGSRVATVELAGATWEVY---ADMDMNYIAY 202
DB 122 ADINHVT--WSGGYELIMILARYGVOPLGSQLATATVEGQTWELMYGVNAGQKTSFVA- 179
OY 203 RRTPTTQSVSELDKAFIDDAVA--RGYIRPEYVL--HAVEGFEIAMEGG-AGLRADPSV 258
DB 180 --ANPTTSF-QGDINDFFKXYLTQNHGFPASSQYLITLITQFGTEPPTGCPATLVNADWSA 236
OY 259 TVQ 261
DB 237 SVQ 239

RESULT 12
US-09-216-295-10
; Sequence 10, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Humicola insolens
US-09-216-295-10

```

```

Query Match 14.8%; Score 213.5; DB 4; Length 253;
Best Local Similarity 30.5%; Pred. No. 1.4e-13;
Matches 73; Conservative 35; Mismatches 99; Indels 33; Gaps 13;

OY 34 EPEPTVELC---GRMDARDVAGGRVYVNNVWGAETA---OCIEV-GLETGNFTITRAD 85
   || || || || || || || || || || || || || || || || || || || || ||
Db 28 EPRQIRSELCXYLYGW-----SGNGYELLNLMNGKKTDTATSGMCTYLDGTNNNGIOWSTAM 82
   || || || || || || || || || || || || || || || || || || || || ||

OY 86 HDNG--NNVAAYPAIYFCGHCACACTNSGLPRVVELSDVR--SWTLPIITGWMNAY 141
   | : || : || : || : || : || : || : || : || : || : || : || : ||
Db 83 EMQAGPVDVVKSPYVKGKQIQG-----RKISDINSMTSVSWTYDR-TDIRANVAY 132
   || || || || || || || || || || || || || || || || || || || || ||

OY 142 DIWEFAPVNSGNGYSGAGELMIMLMNMGSGVMPGSGSRVATVELAGATWEYVADMDMNAYA 201
   || : || : || : || || || || || || || || || || || || || || || ||
Db 133 DVFETARDPDHPR-WGGDYVELMIMTLARVGIYIGTFHSGVNLNAGTWTDM-TGYGNMRY 190
   || || || || || || || || || || || || || || || || || || || || ||

OY 202 YRRITPTTSVSEL--DLKAFITDDAVA-KGYIREPXYLHAVETGFELWEGG-AGLRSADE 256
   | : || : || : || : || : || : || : || : || : || : || : ||
Db 191 YSFLPPSGDIRFSCDIRDFENYLERNHGYPAREONLLIYQVQTECFEGPARFTCRDF 249
   || || || || || || || || || || || || || || || || || || || || ||

RESULT 13
US-08-032-848C-13
; Sequence 13, Application US/08032848C
; Patent No. 5475101
; GENERAL INFORMATION:
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen A.
; APPLICANT: Weiss, Geoffrey L.
; APPLICANT: Larenas, Edward
; APPLICANT: Lorch, Jeffrey D.
; TITLE OF INVENTION: Purification and Molecular Cloning of
; TITLE OF INVENTION: EG III Cellulase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/032.848C
; FILING DATE: MAR 17 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 742-7356
; TELEFAX: 415 742-7217
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 221 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-032-848C-13

Query Match 14.7%; Score 213; DB 1; Length 221;
Best Local Similarity 31.1%; Pred. No. 1.3e-13;
Matches 73; Conservative 40; Mismatches 92; Indels 30; Gaps 12;

OY 40 ELGCGMDARDVAVAGGRVYVNNVWGAET---AOCIEV--GLETGNFTITRADHDNG--NNVA 93
   || : || : || : || : || : || : || : || : || : || : || : ||

```

```

Db      4 QLCDQY--ATYTGYYTINNLMGKDAGSGCCTVNSASSAGTSMSTKMNWMSGGENSVK 61
QY      94 AYPATFGCHMGACTSNGL--PRRYQELSDVRTSWTLPTTGC-RWNAAYDIWSPVT 149
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      62 SY-----ANSGLTFNKKLVSQISQLEPTTARKSYDNTGIRADVAITDLFTADI 108
QY      150 NSNGYSGAELMIWLMNMGVMPGSSRVATVELAGATWEVWY-ADWDNMYIAYRRTPT 208
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      109 NHVT-MSGDELMIMILARYGQVQISQIATATVDGQTWELMWGANGSQKTSFVAPTPI 167
QY      209 TSVSELDKAFIDDAVA-RGYRPEWYLHAVETGFELMEGC-AGLRADTSVYQ 261
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      168 TSE-QGDVNDFFRYLTQNHGFPASSOYLITLQETPEPTGPGATLVSYNSASVQ 221

RESULT 14
US-09-216-295-20
; Sequence 20, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328el Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Emeritella desertoru
US-09-216-295-20

Query Match      14.6%; Score 210.5; DB 4; Length 246;
Best Local Similarity 28.1%; Pred. No. 2.7e-13;
Matches 71; Conservative 33; Mismatches 80; Indels 69; Gaps 12;

QY      40 ELGRDARDVAGGRYRVINNVGAETACIEVLEGTNFTIRADHDNGNNVAAYPAIY 99
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      30 DFCGMDTATV--GNFIVYNLMGQDNA--DSGSQTG-----VDSANGNSIS----- 72
QY      100 FCGHWGACTSNSG-----LPRRYQELSDVRTSWTLPTTGRW----- 137
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      73 ---WHTTWSMSGSSSVKSYANAAVOFTSTKNSLSIPTSW-----KWOYSTTDI 120
QY      138 --NAAYDIWSPYTNNGNGYSGAELMIWLMNMGVMP--GGSRVATVELAGATWEVWY 192
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      121 VANVAITDLF---TSSAGGDSEYEIMIMILALGAGAPISSTGSIATVTLLGCVTWSLYS 176
QY      193 AD---MDWNMYIAYRRTPTTPTTSVSELDKAFIDDAVARGYRPEWYLHAVETGFELMEG- 247
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      177 GPGSMQVYSFVA---SSSTESFSADLMDFIYVLAENGLSSQYLTTHOAGTEPFTGT 232
QY      248 GAGLRADTSVYV 260
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      233 DATLTVSSYSYSV 245

RESULT 15
US-09-216-295-9
; Sequence 9, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328el Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0

```

```

; SEQ ID NO 9
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Humicola grisei
US-09-216-295-9

Query Match      14.6%; Score 210.5; DB 4; Length 253;
Best Local Similarity 28.2%; Pred. No. 2.8e-13;
Matches 70; Conservative 31; Mismatches 96; Indels 51; Gaps 12;

QY      34 EPEPTVELC--GRMDARDVAGGRYRVINNVGAETACIEVLEGTNFTIRADHDNGN 90
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      28 EPRQISICELXYGW-----SGNGYEILNLMGKDIA-----TSGMCCTYLDEGTNNG 74
QY      91 NVAAYPAIYFGCHMGACTSNGLP-----RRRYQELSDVRT--SWTLPTI 132
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      75 -----GIQWNTAMEWQAPDVNKNYPYVGKQIQGRKISDINSMTSVSWTYDR- 123
QY      133 TTGRMNAAYDIWSPYTNNGNGYSGAELMIWLMNMGVMPGSSRVATVELAGATWEVWY 192
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      124 TDLRANAYAVFTARDDPHN-WGQDELMIMILARYGQIPIGTPIHSQVVLAGRITDMLW- 181
QY      193 ADWDNMYIAYRRTPTTPTTSSEL--DKAFITDDAVA-RGYRPEWYLHAVETGFELMEGC- 248
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      182 TGYNGNMRYVSFLPPSGDIDFSCDIDKDFPNYLERNHGYPARQONLIYVOVTECFETGCP 241
QY      249 AGLRSADF 256
        :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      242 ARFTCRDF 249

```

Search completed: August 15, 2002, 10:52:08  
 Job time: 172 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:48:41 ; Search time 29.16 Seconds  
(without alignments)  
860.059 Million cell updates/sec

Title: US-10-003-759-2

Perfect score: 1446  
Sequence: 1 MNMRAVLVLSTLLFGCDM.....PELMGAGLRSADESVTVQ 261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*

1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	352	24.3	382	2	JC2571
2	252.5	17.5	264	2	JU0328
3	213	14.7	237	2	S12610
4	202	14.0	274	2	A72241
5	194	13.4	151	2	H70895
6	163.5	11.3	239	2	S55931
7	153	10.6	258	2	H72240
8	142.5	9.9	332	2	G90291
9	139	9.6	332	2	G90360
10	99	6.8	97	2	H82261
11	97	6.7	673	2	H70528
12	95.5	6.6	322	2	H90425
13	94	6.5	443	2	D72204
14	93	6.4	604	2	B84221
15	93	6.4	741	2	G69514
16	91.5	6.3	690	2	T27357
17	91	6.3	950	2	H87611
18	88.5	6.1	261	1	S12745
19	87.5	6.1	317	2	T36926
20	87	6.0	286	2	S48201
21	87	6.0	1449	1	VG1HFS
22	86.5	6.0	596	2	A55976
23	85	5.9	475	2	T35697
24	85	5.9	497	2	H83886
25	84.5	5.8	198	2	AH2306
26	84.5	5.8	610	2	JH0573
27	84.5	5.8	765	2	S76795
28	84	5.8	468	2	AD0585
29	84	5.8	656	2	H84649

#### ALIGNMENTS

```

30      83      5.7      263      2      S06330      r1cin E - castor b
31      83      5.7      302      2      H90717      hypothetical prote
32      83      5.7      323      2      H85567      hypothetical prote
33      83      5.7      468      2      H64802      yb1m protein - Esc
34      83      5.7      574      2      F75356      serine/threonine p
35      83      5.7      595      2      T29434      beta-galactosidase
36      82.5      5.7      820      2      A40633      chitinase (EC 3.2.
37      82.5      5.7      1090      2      S59077      cellulose 1,4-beta
38      82.5      5.7      1449      1      A43573      E2 glycoprotein pr
39      82      5.7      576      1      RUCSD      r1cin D precursor
40      81.5      5.6      544      2      A47226      dis1-suppressing p
41      81.5      5.6      1374      2      AE3259      extracellular seri
42      81.5      5.6      1447      1      VG1HE2      E2 glycoprotein pr
43      81.5      5.6      1447      1      VG1HE3      E2 glycoprotein pr
44      81      5.6      474      2      G72658      probable Vir B1 A
45      81      5.6      3670      2      T36249      CDA peptide synthe

```

#### RESULT 1

```

JC2571
cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Streptomyces rochei
C:Date: 13-Jun-1995 #sequence,revision 14-Jul-1995 #text,change 22-Oct-1999
C:Accession: JC2571; S34392
C:RefSeq: B1.1-Hanharly-E.; Irdani-T.; Iqbal, M.; McCarthy, A.J.; Mastromeli, G.
Gene 148, 119-124, 1994
A:Title: Characterization and sequence analysis of a Streptomyces rochei A2 endoglucanase
A:Reference number: JC2571; MUID:95011642
A:Accession: JC2571
A:Molecule type: DNA
A:Residues: 1-382 <PER>
A:Cross-references: EMBL:X73953; NID:9393391; PIDN:CAA52139.1; PID:9393392
A:Note: this cellulolytic strain was isolated from the gut of termites
C:Genetics:
A:Gene: egIs
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Superfamily: bacterial cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-37/Domain: signal sequence #status predicted <SIG>
F:38-382/Product: endoglucanase #status predicted <MAT>
F:279-380/Domain: bacterial cellulose-binding domain homology <BCB>
F:280-379/Dissulfide bonds: #status predicted

```

Query Match 24.3% Score 352; DB 2; Length 382;  
Best Local Similarity 36.2%; Pred. No. 1e-21;  
Matches 77; Conservative 32; Mismatches 98; Indels 6; Gaps 4;

```

OY 50 VAGGRRYINNWGAEFTACIEVGLTGNFTTRDHONGNVA--AAPATYFGCHMAC 107
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 49 VIQGRYVNNWNGTSATQCV-TATDSG-FRYTQADGSGVPTNGAKSPVFNCGHYNC 106
OY 108 TNSGILPRVOELSDVPTSWLTPTTGRMNAAYDIPVNSGNGYSGGAELMIWLNW 167
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 107 SNGTLPARISGISAPSISIGFVDNAVYNASDIWIDPPTPTDGV--VNRTEIMWNR 164
OY 168 NGVPGGSRVATVEVLACATWEVYADMDWNIYARRTPPTTSVELDKAFIDAVARG 227
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 165 VQGIQPISSQVGTASVAGRTWVGSGNGCTDVLVSFAPSAMSSWSFVMDVFRATVARG 224
OY 228 YIRPEWYLAHVEVTELMGAGLRSADESVTV 260
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 225 LAGNDWYLTSLQAGEPWNAGLAIVNSFSSTV 257

```

RESULT 2  
JU0328







A:Molecule type: DNA  
A:Residues: 1-673 <COL>  
A:Cross-references: GB:296796; GB:AL123456; NID:g3261797; PIDN:CAB09571.1; PID:g3261799  
A:Experimental source: strain H37Hv  
C:Genetics:  
A:Gene: RV0457c  
C:Superfamily: prolyl oligopeptidase

Query Match 6.7%; Score 97; DB 2; Length 673;  
Best Local Similarity 22.7%; Pred. No. 2.7;  
Matches 59; Conservative 16; Mismatches 67; Indels 118; Gaps 13;

QY 32 EPEPEPVE-LCGR-WDARDVAGGRYVNNVWGAETPAQCTEVLGTFGNTTTRADHNG 89  
DB 455 ETSRTPYDVGLGRILMLAR--GCTYALANIRGGGE----- 487  
QY 90 NNVAAYPAIYFGCHWGAAGTNSGLPRVQELSDVTSWTLPTTGRNNAAYDIWFSPT 149  
DB 488 -----YGPCHWQAMREGDKVAQDFAAVATDLVTRGCTTA----- 523  
QY 150 NSGNGSGAGELMTIMLWNGVMPG-----GSRVATVE-----LAGATWE 189  
DB 524 -EQGARGGS-----NGGLLMGIMLTGYPEKFGALYCDVPLDMKRYHLLLAGASGM 574  
QY 190 VWYAD-----MDWNTI-----AYRRTT---TTSVSLDLKAFIDDAVARGYIR-- 230  
DB 575 AEYDPPNDPMKRFISEXPYQNTISANKRPPVLTMTSTR-----DDRHHGHAHRM 626  
QY 231 -----PEMYLHAIVETG 241  
DB 627 TAAQAAGHPVWYENIEGCG 646

## RESULT 12

H90425  
hypothetical protein celB [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus  
C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: H90425  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Adayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
arrett, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.  
submitted to Genbank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A9139  
A:Accession: H90425  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-322 <KUR>  
A:Cross-references: GB:AE006641; NID:g13815835; PIDN:AAK4263.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: celB

Query Match 6.6%; Score 95.5; DB 2; Length 322;  
Best Local Similarity 20.3%; Pred. No. 1.6;  
Matches 44; Conservative 33; Mismatches 95; Indels 45; Gaps 8;

QY 72 VGLTEGNFTIT-----RADHDGNNVAAYPAIYFG--CHWGACTSNS----- 111  
DB 83 IGYALGNVNMNTINITYLVAINALSKISSNVVDGTYGGLMAGGLMPEMRTQLOFL 142  
QY 112 GLPRVQELSDVTSWTLPTTGRNNAAYDIWFSPTNSGNGSGAGELMTIMLWNG 169  
DB 143 SLPMYVLRPLNPNYSILNVSYLINGSIDFSYDMLSONPNITSLOYGDFEIMIMWYNE 202  
QY 170 G-----VMGGSNVAIV--ELAGATWEVYA-----DMDNNTIAYKRTPPTTSYS 212  
DB 203 NLSHNPFYIYGNMSIPLINKINKINLSMEVYVLRPTGSANQMTGVYFLSPLKEKAEFG 262  
QY 213 E-----LDLKAFLDDAVARGYIRPEWYLHAIVETGFE 243

DB 263 VPIGYILKNNMSYIEKAGVNIYNNFTYLDALQYGM 299

## RESULT 13

D72204

hypothetical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: D72204  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic-  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72204; MUID:99287316  
A:Accession: D72204  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-443 <ARN>  
A:Cross-references: GB:AE001821; GB:AE000512; NID:g4982421; PIDN:AAD36903.1; PID:g498  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1841

Query Match 6.5%; Score 94; DB 2; Length 443;  
Best Local Similarity 23.2%; Pred. No. 3;  
Matches 54; Conservative 21; Mismatches 80; Indels 78; Gaps 13;

QY 1 MNVRAVLVLSLLIFGCDWL-----FPDGD-----NGKEPEPEPEYELCGRDARDA 51  
DB 172 VNQYKAVFAISGGTGGGQWQVDFPNQPDHDSAEENGIGPPAPVNVFVC-----DKE 225  
QY 52 GGRYRVNNWGAE--TAQCTEVLGTFGNF-----TTRADHNGNNV--AAVPAIYF 100  
DB 226 GNVSEIINTSEPEQPTESVYIVDGLSDFGTPVATSTNPGGNGANLYRYTYDAYTL 285  
QY 101 GCHWGACTNSGGLPRVQELSDVTSWTLPTTGRNNAAYDIWFSPTNSGNGSGAGAE 160  
DB 286 --YIGFDIQN-----GSMVYAIIG--GIDTRPSGYTGDS 317  
QY 161 LMIWLNWNGVMPGSGRVAIVELAGATWEVYA-----DWNV 199  
DB 318 A-----WGRKISFGADYAVDEI-----YFWWAGSGGLDSNNFCENGLDMNY 360

## RESULT 14

B84221  
hypothetical protein Vng0631c [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84221  
R:Ng, W.V.; Kennedy, S.P.; Mahliras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky  
J.; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja-  
Jung, K.H.; Alam, M.; Freltas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebbhardt, H.; Lowe, T.M.;  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: B84221  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-604 <STO>  
A:Cross-references: GB:AE004437; NID:g10580222; PIDN:AA919134.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG0631C

Query Match 6.4%; Score 93; DB 2; Length 604;  
Best Local Similarity 23.3%; Pred. No. 5.2;  
Matches 59; Conservative 27; Mismatches 93; Indels 74; Gaps 13;

QY 3 VNRATLVISLILIFGCDMLFPDGNKKEPEPEP-----EPTVELCGRDARDVAGG 53

```
Db 107 VARAPVAAGGLPGVAVLF-----RGRLDDEEVIALAGLAFPTLVYYSRFRMSDYLVA 162
QY 54 RYRYINNVWGAETAGCLEVEGETETGNTTTRADHDNG-----NNVAA 94
Db 163 AF-----SLAAGFAV-RA-HDTGRMILLPVAAGMLALATLAKENALV 203
QY 95 YPAIFEGCHWGACTNSGLPRVOELSDVRSWTLPITTGKMAAYDIWFSPVYNSGNG 154
Db 204 YAMFAGAG--GALVAD-----RLLITANPRGLSMTOS--LQAGVTRRAARGFWAMRRTLAASA 256
QY 155 YSGAELMIMLNK---GCY-----MPGSSRAVVELAGATWEVYADW---DMNYIAY 202
Db 257 VVAGAVFAVFAETPRPVAGGIGAAPTRLPSSAVAGSADAHAHALMGTWVVDGVDHSTIAY 316
QY 203 ---RRTPTTSV 211
Db 317 LVVALRTLATTAIV 329
```

## RESULT 15

G69514

hypothetical protein AF2119 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Nov-1999

C:Accession: G69514

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Ueberlack, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; M01D:98049343

A:Accession: G69514

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-741 &lt;LE&gt;

A:Cross-references: GB:AE000958; GB:AE000782; NID:92689281; PIDN:AA89151.1; PID:9264842

C:superfamily: Archaeoglobus fulgidus hypothetical protein AF2119

Query Match 6.4% Score 93; DB 2; Length 741;  
Best Local Similarity 24.4%; Pred.No. 6.5;  
Matches 49; Conservative 21; Mismatches 51; Indels 80; Gaps 11;

```
QY 75 ETGNFTITRADHDNGNNVAAYPA-----IYFGCHWGACTNSG-----LPRR-- 116
Db 204 DTDITTLVAKVYNSGNPNVDYVKFAEPDSQRIELG---GAMTNSSGIAKLSFIPKNVG 260
QY 117 -----VOELSDVRSWTLPITTGKMAAYDIWFSP---VTNSGN-GYSGAELM 162
Db 261 LSDKIRVNFVAKIEDVMTNCNAYTTTNRAILAEVVAITPGSYDITLVGRMYTSGGAD-W 319
QY 163 IWLNV---NGVMP-----GGSRAVVELAGATWEVY----- 192
Db 320 VRVNVYVDENVSLKPIPVVQKFTGNR-----ASVTIMKYGLDDYCTDPNCHREGI 371
QY 193 -----ADMDMNYIAYRRTTP 207
Db 372 YGNFDADWDGACIAYGSTTP 392
```

Search completed: August 15, 2002, 10:51:39  
Job time: 178 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:51:42 ; Search time 15.93 Seconds

(without alignments)  
634.388 Million cell updates/sec

Title: US-10-003-759-2

Perfect score: 1446  
Sequence: 1 MNMRAVLVLSLLIFGCDW.....FELMEGAGLSADEFSVTWQ 261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	252.5	17.5	264	1 GUNS_ERMCA	P16630 erwilia car
2	213	14.7	237	1 GUN_ASPAC	P22659 aspergillus
3	163.5	11.3	239	1 GUN_ASPAC	Q12679 aspergillus
4	93	6.4	625	1 BGAL_LACSK	Q48846 lactobacill
5	93	6.4	741	1 YL19_ARCFU	Q28161 archaeoglob
6	88.5	6.1	261	1 XYNA_CLOSA	P7137 clostridium
7	88	6.1	269	1 EL2_BOVIN	Q29461 bos taurus
8	87	6.0	286	1 GUB_RHOMR	P45788 rhodothermu
9	87	6.0	1449	1 VGL2_CVPEF	P18450 porcine tra
10	85	5.9	475	1 ABFB_STRCO	O54161 streptomyce
11	84.5	5.8	610	1 CHIT_STRPL	P11220 streptomyce
12	84	5.8	1447	1 VGL2_CVPEP	Q02167 porcine tra
13	83	5.7	468	1 YBFM_ECOLI	P75737 escherichia
14	83	5.7	475	1 ABFB_STRPL	P96463 streptomyce
15	82.5	5.7	1090	1 GUXB_CELFI	P50899 cellulomona
16	82.5	5.7	1449	1 VGL2_CVPEP	P33470 porcine tra
17	82	5.7	576	1 RIC1_RICCO	P02879 ficinus com
18	81.5	5.6	224	1 GPH_PASMU	O9CK15 pasteurella
19	81.5	5.6	225	1 TTRR_CORGL	P71121 corynebacte
20	81.5	5.6	544	1 DSK1_SCHPO	P36616 schizosach
21	81.5	5.6	1447	1 VGL2_CVPEP	P07946 porcine tra
22	80.5	5.6	318	1 ALYS_BPHB3	P32762 streptococ
23	80.5	5.6	376	1 SERC_MYCTU	O10534 mycobacteri
24	80	5.5	5038	1 RYR1_HUMAN	P21817 homo sapien
25	79.5	5.5	700	1 YNCD_ECOLI	P76115 escherichia
26	79.5	5.5	820	1 CHIA_ALTOS	P32853 alteromonas
27	79.5	5.5	986	1 GUN2_CLOS	P23659 clostridium
28	79	5.5	2334	1 WAPA_BACSU	Q07833 bacillus su
29	78.5	5.4	233	1 XYNA_MAGGR	P55335 magnaporthe
30	78.5	5.4	479	1 CBSA_SULSO	P58029 sulfolobus
31	78.5	5.4	966	1 ENV_CAEYC	P16266 caprine art
32	78.5	5.4	1447	1 VGL2_CVPEP	Q01977 porcine tra
33	78	5.4	509	1 YAGE_SCHPO	Q9ut44 schizosach

34	78	5.4	852	1 POL_BLVJ	P03361 bovine leuk
35	77.5	5.4	826	1 ACMI_DROME	P16395 drosophila
36	77.5	5.4	744	1 PAC_ECOLI	P06875 escherichia
37	77.5	5.4	877	1 AGLI_HORVU	Q43763 hordeum vul
38	77.5	5.4	1356	1 HET1_PODAN	Q00808 podospora a
39	77	5.3	572	1 SYM_AERPE	Q9CY33 aeropyrum p
40	77	5.3	1481	1 APV_THEET	P38339 t amylopull
41	77	5.3	2044	1 SIF2_DROME	P16320 drosophila
42	76.5	5.3	216	1 SPRB_IPOBA	P10965 ipomoea bat
43	76.5	5.3	270	1 KITM_MOUSE	Q9I088 mus musculu
44	76.5	5.3	318	1 ALYS_STRPN	P06553 streptococ
45	76.5	5.3	327	1 A85B_MYCLE	P31951 mycobacteri

## ALIGNMENTS

```

RESULT 1
ID GUNS_ERMCA STANDARD; PRT; 264 AA.
AC P16630;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase S precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase S)
DE (Cellulase S).
GN CELS.
OS Erwilia carotovora.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=554;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=SCC3193;
RA MEDLINE=90337352; PubMed=2379837;
RA Saadilabti H.T., Henrissat B., Palva E.T.;
RT "Cels": a novel endoglucanase identified from Erwilia carotovora
RT subsp. carotovora.";
RL Gene 90:9-14(1990).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC -1- linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY H (FAMILY 12 OF GLYCOSYL
CC HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M33399; AAA24817.1; -
CC PIR: J00328;
CC DR InterPro: IPR002594; Glyco_hydro_12.
CC DR Pfam: PF01670; Glyco_hydro_12; 1.
CC DR ProDom: PD004316; Glyco_hydro_12; 1.
CC KW Cellulose degradation; Hydrolase; Glycosidase; signal.
CC FT SIGNAL 1 32
CC FT CHAIN 33 264
CC FT SEQUENCE 264 AA; 29757 MW; E6D61388950C77AA CRC64;
SO

```

Query Match 17.5%; Score 252.5; DB 1; Length 264;  
Best local similarity 31.6%; Pred. No. 1,4e-14;  
Matches 72; Conservative 39; Mismatches 86; Indels 31; Gaps 10;

```

OY 54 RYRYINNVGAETRAQ-----CIEVGLTNGFTITRADHDGNNVAAYPAITFGCH 103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 48 KYYLENNVWGEDEIKGQQTIFYNSPIISMG---WNHWPSTH---SVKAYSLVSGWH 100
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 104 WGA-CTNSNGLPKRVQGLSVRTSWTLPLITTGWNNAYDIWFSPVATNSGNGVSGAEIM 162
      | | | | | | | | : : : : : : : : : : : : : : : : : : : :

```



Db 19 MCSQYDS--ASSPYYVNQNIAMGEYGTGSCQYVVDKLSSSGASMHKRTWMSGGEYIVKS 76

QY 95 YPAIFYGCHMGACHTSNGL---PRVOELSDVRTSWTLPTITTG--RNAAYADIFESPVTN 150

Db 77 Y-----SNSLTFPKDKLTVSSVSIPTSVYMSQDDPTNVQADVSDL-FTANANA 122

QY 151 SGNGISGCAELIMILNMNGVWPGGSRVAYVELAGATWEVY----ADMDWNIYART 205

Db 123 DHATSSGCEYELIMILNARVGSVQPIGKQIATVATVAGKSMEEVYGTSTQAGAEQKTYSPVAG 182

QY 206 TPTTSVSELDKAFIDDAVA-RGIRREPWYIHAVETGFELMEGAGLRSD 255

Db 183 SPINSMWS-DIKDFENYLTQNGEPASSOHLITLQCGTEPTGPATFTVD 232

```

RESULT      4
BGAL_LACSK STANDARD:          PRT:       625 AA.
AC Q48B46;
DT 15-DEC-1998 (Rel. 37, Created)
DPT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
CN Beta-galactosidase large subunit (EC 3.2.1.23) (lactase).
LACT.
OC Lactobacillus sakei.
OC Bacteria; Firmicutes; Bacilllus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
RX NCBI_TaxID=1599;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 20017;
RL MEDLINE=96118231; PubMed=8574399;
RA Obst M., Meding E.R., Vogel R.F., Hammes W.P.:
RT "Two genes encoding the beta-galactosidase of Lactobacillus sake.",
RLE Microbiology 141:3059-3066(1995).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
    galactose residues in beta-D-galactosides.
CC -I- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.
CC -----
CC -I- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between The Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/commercial/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X82287; CAAS7730.1; -.
DR DRSSP; P00722; IBGL.
DR InterPro: IPR001649; Glyco_hydro_2.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2.C; 1.
DR Pfam; PF02837; Glyco_hdro_2.N; 1.
DR PRINTS; PR00132; GLYHYDLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2.1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2.2; 1.
KW Hydrolase; Glycosidase.
FT ACCT_SITE 533 533         PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 465 465         NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE   625 AA; 72457 MW; C658A3CA613CB886 CRC64;.
Query Match           6.4% Score 93; DB 1; Length 625;
Best Local Similarity 17.9%; Pred. NO. 1.5;
Matches 44; Conservative 39; Mismatches 87; Indels 76; Gaps 11;
```

```

Db 399 ---DQAGIYMAAETINLESHGSMQKMGAVEBWNPGSIDENEHAATLDRATINETPEKNHY 455
QY 160 ELMTIWLNNNGCGVMPGGSRHATVELAGATWEVYADMDMNYIAVRRTPPT-----208
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 456 SILFP-----SLGNESTAGSVLEBKMN-----YYQQDPTPLVHYEGVFR 466
QY 209 ----TSVSEL-----DIKAFIDDAVARGYRPEMYLHAIVETG-----F 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 497 PEYKATISDVESRMVAFPAIKAVLIDNAPQKPFILCE-YHMDGMSNLSGMSQSIDLLSQ 555
QY 243 ELMBEG 248
   : : : : |
Db 556 DMYOG 561

```

Query Match	6.4%	Score 93	DB 1	Length 741
Best Local Similarity	24.4%	Pred. No. 1.8		
Matches 49	Conservative 21	Mismatches 51	Indels 80	Gaps 11

**OY** 53 GRRVYNV-----NGAETAOCEVLEGLETGNTTTPADHDNNNAAP--AIYEGCHWG 107  
:  
:  
:  
**Dd** 342 GKRLIVGVNHEHMPETGRITTADEAMDIACQNRHNINAVRSHYPDRLSFVNGC--- 396  
:  
:  
:  
**OY** 106 ACTNSGLPRVQELSDVRITSWTLTPITGGMN--AAVDIFSEPVTVSG----NGYSGGA 155  
:  
:  
:

```

0Y 75 ETGNETITRADHMGNNVAAPA-----TYFGCHWACTSNSG-----LPRR- 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 204 DDDITLVAKVYNSGNPPDPYKVFYAEFDQSRIFLG---GAMNSSGIKLSPIRNVG 260
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 117 -----VOELSDVRTSWTLPTTIGRMAAYDIMEFP-----VTNSGN-GYSGGAELM 162
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```





```

FT CHAIN 17 1449 E2 GLYCOPROTEIN.
FT DOMAIN 17 1390 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1391 1410 POTENTIAL.
FT DOMAIN 1411 1449 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1411 1432 CVS-RICH.
FT CARBOHYD 26 26 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 516 516 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 704 704 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 819 819 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 840 840 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1200 1200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1294 1294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1311 1311 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1324 1324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1358 1358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1449 AA; 159957 MW; 971BBAE191FDIAF CRC64;

```

```

Query Match 6.0%; Score 87; DB 1; Length 1449;
Best Local Similarity 23.0%; Pred. No. 12;
Matches 45; Conservative 20; Mismatches 69; Indels 62; Gaps 10;

```

```

QY 45 WDARVAGGRVYNNVNGAETACIEVLEGTGNTTRADHNGNNV-AAVPAIFYG-- 101
DB 88 WD-----YATENSTWNHK--QRLNVVNGVYSITVTTTRNNSAGATICKGSP 137
QY 102 -----CHMGA-----CTNSGLPRVQELSDVRTSWTLPTTGGWMAAY 141
DB 138 PTTTSSLTCNMGSECLNHNKFPICPSNS-----EANGCNMLY 176
QY 142 DI-WESPVTNSGNGSGAELMIWL--NNGGVMGPGSRVATVELAGATWIEWYWD--WD 196
DB 177 GLQNFADAIVA---YLHGSYRISFENQMSGTVTLGDMRATTLTAGTLVLDLWENPYVD 233
QY 197 WNYIAYRTTPTTSVS 212
DB 234 VSYRVNKNKGTIVYS 249

```

```

RESULT 10
ID ABFB_STRCO STANDARD; PRT; 475 AA.
AC 054161;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-L-arabinofuranosidase precursor (EC 3.2.1.55) (Arabinosidase).
GN ABFB OR SC7H1.02.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

```

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Murphy L., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
arabinofuranoside residues in alpha-L-arabinosides.
CC -1- PATHWAY: XYLAN DEGRADATION.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 62 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
or send an email to license@sdb.ch).
CC -----
CC EMBL: AL021411; CAAL6189.1; -.
CC InterPro: IPR000772; Ricin_B_lectin.
CC Pfam: PF00652; Ricin_B_lectin; 1.
CC SMART: SM00458; Ricin; 1.
CC PROSITE: PS50231; RICIN_B_LECTIN; 1.
CC Xylan degradation; Hydrolase; Glycosidase; Signal; Lactin.
FT SIGNAL 1 37
FT CHAIN 38 475 ALPHA-L-ARABINOFURANOSIDASE.
FT DOMAIN 39 166 RICIN B-TYPE LECTIN.
SQ SEQUENCE 475 AA; 50045 MW; 47E07EE543CA60D CRC64;

```

```

Query Match 5.9%; Score 85; DB 1; Length 475;
Best Local Similarity 21.8%; Pred. No. 5.2;
Matches 53; Conservative 31; Mismatches 93; Indels 66; Gaps 14;

```

```

QY 11 SLLLFQGC-----DWLPFD-----GDN-----GKEPEPEPEVELCGWRDARDVAG 52
DB 63 ALLQLYDCWGTNQOWTSTDTGRITVYGGDKLDVGHATAPGTRVOI-----WSCSGGAN 117
QY 53 GRYRVYNN--VWGAETACIE-VGLETGNTTRADHNG-----NNVAA 94
DB 118 QQWRVNSDGTIVGVESGLCELAAGAGTAGTAVQMLTCCGNGQWGTGLTGPPTDGCIA 177
QY 95 YPAITFGCHMGACTNSGLPRVQELSDVRTSW-TLPITTTGRWMAAYDIWESPVTNSGN 153
DB 178 LPSTY---RW-----SSTGV-----LAQPKSGWVALKDFTTYTHNGRHLVYGS--TSSGS 222
QY 154 GYSGGAELMIWNNMGVMPGSGRATVELAGA-----TWYWTADMDMNYIAYRRT 205
DB 223 SY-GSMVFSPTNMSDMSASQANNAQAAVAPTLTFYFAPKNIWLYLAWGSMPIFYRTSS 281
QY 206 TPT 208
DB 282 DPT 284

```

```

RESULT 11
ID CHIT_STRPL STANDARD; PRT; 610 AA.
AC P11220;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Chitinase 63 precursor (EC 3.2.1.14).
GN CHTA.
OS Streptomyces plicatus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1922;
RN [1]

```



FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 832 832 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 838 838 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1072 1072 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1198 1198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1292 1292 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1309 1309 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1322 1322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1334 1334 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1339 1339 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1356 1356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1369 1369 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 1447 AA; 160115 MW; 062EBD052DE2637 CRC64;

Query Match 5.8%; Score 84; DB 1; Length 1447;  
 Best Local Similarity 23.4%; Pred. No. 22;  
 Matches 51; Conservative 24; Mismatches 73; Indels 70; Gaps 12;

QY 37 PYVE---LCGRDADYAGGRVNNV---WGAET-----AACIEVLETGNETT 82  
 DB 60 PIVQPFNCRINDSNDL---YVLENLKALYWDATENTNMHRLNRYVNGYRST 115  
 QY 83 RADHDGNNV-AAYPAIYFG-----CHMGA-----CTNSGLPRVOE 119  
 DB 116 VITTRFNENAEGLIICICGSPPTTSSSLTCNMGSECLRNKRPICPSNS----- 167  
 QY 120 LSDVTSWTLPTITGRNMAAYDI-WFSPVNSGNGYSGAELMIL--NMNGVMPGGS 176  
 DB 168 -----EANGCMMLYGLWFA---DEVAYILHGAISYRISFENMGSGVTFEGDM 211  
 QY 177 RYATVEIAGATWEVYAD--WDMNTIATRTPTTSVS 212  
 DB 212 RYATLEVAGTLVDLWFMFNPYDYVSYRVNKNKGTIVS 249

RESULT 13  
 YBEM\_ECOLI STANDARD: PRT; 468 AA.  
 AC P75733;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein ybfm.  
 GN YBEM OR B0681.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=9742617; PubMed=9278503;  
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RT Science 277:1453-1474(1997).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-K12;  
 RC MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiiuchi T.;  
 RT "A 718-bp DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map.";

RL DNA Res. 3:137-155(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AEO00172; AAC73775.1;  
 DR EMBL; D90707; BAA35329.1; ALT\_INIT.  
 DR Ecogene; EG13659; ybfm.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 468 AA; 52780 MW; 954B5A778A61C2E4 CRC64;

Query Match 5.7%; Score 83; DB 1; Length 468;  
 Best Local Similarity 23.6%; Pred. No. 7.5;  
 Matches 46; Conservative 25; Mismatches 70; Indels 54; Gaps 10;

QY 83 RADHDGNNVAYPAIYFCHMGACTSNGSLPRVOELSDVTSWT----- 128  
 DB 281 RSYNDLYDGTAMLDALTFG-----YRADVDLRLGCTWAKDGGGYFLQR 327  
 QY 129 LPTITGRNMAAYDIWFSPVNSGNGYSGAELMIL-----NMNGVMPGGRVATV 181  
 DB 328 MTP-TYASSNGRLDMW---DNRSDFNNGEKAFFGAYDLKMN---LPGFAGASY 379  
 QY 182 ELAGATWEVYADWDMNTIAY--RRTPTTSVSELDLKAIFIDAVARGYIREVWLHVE 239  
 DB 380 VYA---WDKAPATWQSNPDYDKNRTIESAVSLDAVYTTDGNRAKGM---FKLHFE 433  
 QY 240 ----TGFELMEGAG 250  
 DB 434 YDNHSDIPSMGGYG 448

RESULT 14  
 ABFB\_STRLI STANDARD: PRT; 475 AA.  
 ID ABFB\_STRLI  
 AC P96463;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alpha-L-arabinofuranosidase precursor (EC 3.2.1.55) (Arabinosidase).  
 GN ABFB.  
 OS Streptomyces lividans.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1916;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=66 / 1326;  
 RX MEDLINE=97220396; PubMed=9148759;  
 RA Vincent P., Shareck F., Dupont C., Morosoli R., Kluepfel D.;  
 RT "New alpha-L-arabinofuranosidase produced by Streptomyces lividans:  
 RT cloning and DNA sequence of the abrf gene and characterization of the  
 RT enzyme";  
 RL Biochem. J. 322:845-852(1997).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=1326;  
 RA Shareck F.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 CC  
 CC - FUNCTION: HAS A SPECIFIC ARABINOFURANOSE-DEBRANCHING ACTIVITY ON  
 CC XYLAN FROM GRAMINAE. ACTS SYNERGISTICALLY WITH THE XYLANSSES AND  
 CC BINDS SPECIFICALLY TO XYLAN. FROM SMALL ARABINOXYLO-OLIGOSIDES,  
 CC IT LIBERATES ARABINOSE AND, AFTER PROLONGED INCUBATION, THE  
 CC PURIFIED ENZYME EXHIBITS SOME XYLANOXYLYTIC ACTIVITY AS WELL.  
 CC - CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-  
 CC arabinofuranoside residues in alpha-L-arabinosides.  
 CC - PATHWAY: XYLAN DEGRADATION.



CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 62 OF GLYCOSYL HYDROLASES.  
 CC -1- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M64551; AAC26524.1; -  
 DR InterPro: IPR000772; Ricin\_B\_lectin.  
 DR Pfam: PF00652; Ricin\_B\_lectin; 1.  
 DR SMART: SM00458; RICIN; 1.  
 DR PROSITE: PS50231; RICIN\_B\_LECTIN; 1.  
 KW Xylan degradation; Hydrolase; Glycosidase; Signal; Lactin.  
 FT SIGNAL 1 37  
 FT CHAIN 38 475 ALPHA-L-ARABINOFURANOSIDASE.  
 FT DOMAIN 39 166 RICIN B-TYPE LECTIN.  
 FT SEQUENCE 475 AA; 50369 MW; C3CB14EE7BF85AD CRC64;  
 SO  
 Query Match 5.7%; Score 83; DB 1; Length 475;  
 Best Local Similarity 22.4%; Pred. No. 7.7;  
 Matches 55; Conservative 31; Mismatches 88; Indels 72; Gaps 15;  
 OY 11 SLLEFGC-----DMLPPD-----GDN-----GKEPEPEPTVELCGRWADRVAG 52  
 DB 63 ALLQLIDWGTGNTQOWTSTDTGRLTVGDKLDVPGHATPCTRQIWSG-----SG 114  
 OY 53 GR--YRVINN--VWGAEATACIE-VGLETFNITRADHDNG-----NN 91  
 DB 115 GRNQMRVNSDGTGVVSGCLTEAAGAGTPGTAQLMTGCGGNQKWTGLTGPPTDG 174  
 OY 92 VAATYATIFGCHMGACTNSGLPRVQELSDVRTSM-ILTPITTRKMAADYIMSPYTN 150  
 DB 175 TCALPSTY---RW-----SSTGV-----LAOPKSGWALKDETTVTHNGRLHYGS--TS 219  
 OY 151 SGNXSGCAELIMLMNMGVMPGSRVATVELAGA-----TWYVYADMDNNYAY 202  
 DB 220 SSSST--GSMVTSPTFNMSDMASAGGNANQAAVATLTFYFAPKNIVLWLYGMSPPFYR 278  
 OY 203 RRTPT 208  
 DB 279 TSSDPT 284  
 RESULT 15  
 GUXB\_CELFI STANDARD; PRT; 1090 AA.  
 AC P50899;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellulohydrolase B)  
 DE (1,4-beta-cellulohydrolase B) (CBP120).  
 GN CBHB OR CENE.  
 OS Cellulomonas fimi.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.  
 OX NCBI\_TaxID=1708;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.  
 RC STRAIN=ATCC 484;  
 RX MEDLINE=96003898; PubMed=7575482;  
 RA Shen H., Giles N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;  
 RT "Cellulohydrolase B, a second exo-cellulohydrolase from the  
 RL "Cellulolytic bacterium Cellulomonas fimi.";  
 RL Biochem. J. 311:67-74(1995).  
 RN [2]  
 RP SEQUENCE OF 54-75.

RX MEDLINE=93209933; PubMed=8458833;  
 RA Melnik A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;  
 RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase  
 RT D (Cend), a family A beta-1,4-glucanase.";  
 RL J. Bacteriol. 175:1910-1918(1993).  
 RN [3]  
 RP SEQUENCE OF 54-78.  
 RX MEDLINE=94197708; PubMed=8147863;  
 RA Shen H., Tomme P., Melnik A., Gilkes N.R., Kilburn D.G.,  
 RA Warren R.A.J., Miller R.C. Jr.;  
 RT "Stereochemical course of hydrolysis catalysed by Cellulomonas fimi  
 RT Cend, a member of a new family of beta-1,4-glucanases.";  
 RL Biochem. Biophys. Res. Commun. 199:1223-1228(1994).  
 CC -1- FUNCTION: HYDROLYSE CELLULOSE TO A MIXTURE OF CELLOTRIOSE,  
 CC CELLOTRIOSE AND CELLOBIOSE, WITH ONLY A TRACE OF GLUCOSE. IT  
 CC HYDROLYSED CELLOPENTAPOSE TO CELLOTRIOSE AND CELLOBIOSE, AND  
 CC CELLOTRIOSE TO CELLOBIOSE, BUT IT DID NOT HYDROLYSE CELLOTRIOSE.  
 CC HAS ALSO WEAK ENDOLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS  
 CC WITH INVERSION OF ANOMERIC CONFIGURATION.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
 CC in cellulose and celloetraose, releasing cellobiose from the non-  
 CC reducing ends of the chains.  
 CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN  
 CC (CBD).  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: L38827; AAB00822.1; -  
 DR HSSP: P07986; IEXG.  
 DR InterPro: IPR001919; CBD\_2.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003962; FNIII\_repeat.  
 DR InterPro: IPR000556; Glyco\_hydro\_48.  
 DR Pfam: PF00553; CBD\_2; 1.  
 DR Pfam: PF00041; fn3; 3.  
 DR Pfam: PF02011; Glyco\_hydro\_48; 1.  
 DR PRINTS: PR00014; FNTYPERIII.  
 DR PRINTS: PR00844; GLHYDRLASE48.  
 DR PRODOM: PD011903; Glyco\_hydro\_48; 1.  
 DR SMART: SM00060; FN3; 3.  
 DR PROSITE: PS00561; CBD\_BACTERIAL; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 FT SIGNAL 1 33  
 FT PROPEP 34 53  
 FT CHAIN 54 1090  
 FT DOMAIN 54 699 EXOGLUCANASE B.  
 FT DOMAIN 700 785 CATALYTIC (BY SIMILARITY).  
 FT DOMAIN 794 884 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 891 978 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 989 1090 FIBRONECTIN TYPE-III 3.  
 FT ACT\_SITE 513 513 CELLULOSE-BINDING (BY SIMILARITY).  
 FT DISULFID 990 1089 NUCLEOPHILE (BY SIMILARITY).  
 FT SEQUENCE 1090 AA; 114829 MW; 046BB9D956F2F399 CRC64;  
 SO  
 Query Match 5.7%; Score 82.5; DB 1; Length 1090;  
 Best Local Similarity 23.8%; Pred. No. 21;  
 Matches 41; Conservative 12; Mismatches 52; Indels 67; Gaps 8;  
 OY 45 WDARDVAGR-----YRVINNWGAETACIEVGLGTGN-FTTRADHDNGNNVAAP 96  
 DB 915 WNASDTGSGSLKGYDYVAGATRVGSTTASVTDGLPAATAYQYTVATDAGNVSA-- 972  
 OY 97 ATYFGCHMGACTNSGLPRVQELSDVRTSMILTPITTRKMAADI--WFS-----PV 148

Db 973 -----ASAALSVTTKTPOTGGSCSVAYXNASMNSGFTASVRI 1009  
OY 149 TNSG---NGYSGAEL-----MTWLN-----WNGGVMPG 174  
Db 1010 TNGTTTNGWSLGFDLTAGOKVQCGWSATWTQSGSTVTATNAPWNGTLPAG 1061

Search completed: August 15, 2002, 10:58:02  
Job time: 380 sec



```

Db      61 VKAETACIEVEGTEGNTTTTRADHDNGNNVAAYPAITGCHMAPRAIRDCARAGAV 120
QY      115 RVOELSDVTSWLTPTTGRNNAAYDIWSPVTSNGSGGAEIMILMNGGVPMP 174
Db      121 RRAHELD-----VPTTGRNNAAYDIWSPVTSNGSGGAEIMILMNGGVPMP 173
QY      175 GSAVATVELAGATWVWYADMDNYIAYRRTPPTSVELDKAFIDAVARGIPEWY 234
Db      174 GSAVATVELAGATWVWYADMDNYIAYRRTPPTSVELDKAFIDAVARGIPEWY 233
QY      235 LHAETGFELEGGAGLSADFSVTYV 261
Db      234 LHAETGFELEGGAGLSADFSVTYV 260

RESULT 2
ID      09K1H1      PRELIMINARY;      PRT;      371 AA.
AC      09K1H1:
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      CELULIASE 12A.
GN      CEL12A.
OS      Streptomyces sp. 11A68.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX      NCBI_Taxid=133452;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=11A68;
RA      van Solingen P., Meijer D., van der Kleij W.A.H., Barnett C.C.,
RA      Bolle R., Power S.D., Jones B.E.;
RT      "Cloning and expression of an endocellulase gene from a novel
RT      Streptomyces isolate from an East African soda lake.";
RL      Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR      EMBL: AF233376; AAF91283.1; -.
DR      HSSP: P07986; 1EXG.
DR      InterPro: IPR002594; Glyco_hydro_12.
DR      InterPro: IPR001230; Prenyltn.
DR      Pfam: PF01670; Glyco_hydro_12; 1.
DR      ProDom: PD004316; Glyco_hydro_12; 1.
DR      ProSITE: PS00294; PRENYLATION; UNKNOWN_1.
SQ      SEQUENCE 371 AA: 38481 MW: 0E1BC4288A148914 CRC64;

Query Match      27.4%; Score 396.5; DB 2; Length 371;
Best Local Similarity 39.7%; Pred. No. 6.7e-24;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

```

```

DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      CELULIASE (EC 3.2.1.4).
GN      EGLS.
OS      Streptomyces rochei (Streptomyces parvulus).
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX      NCBI_Taxid=1928;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=A2;
RX      MEDLINE=95011642; PubMed=7523249;
RX      Perito B., Hanhart E., Irdani T., Iqbal M., McCarthy A.J.,
RA      Mastromel G.;
RT      "Characterization and sequence analysis of a Streptomyces rochei A2.";
RL      Gene 148:119-124(1994).
DR      EMBL: X73953; CA452139.1; -.
DR      HSSP: P07986; 1EXG.
DR      InterPro: IPR001919; CBD_2.
DR      InterPro: IPR002594; Glyco_hydro_12.
DR      Pfam: PF00553; CBD_2; 1.
DR      Pfam: PF01670; Glyco_hydro_12; 1.
DR      ProDom: PD004316; Glyco_hydro_12; 1.
DR      ProSITE: PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW      Hydrolase; Glycosidase.
SQ      SEQUENCE 382 AA: 39398 MW: 21C014342EFC6565 CRC64;

```

```

Query Match      24.3%; Score 352; DB 2; Length 382;
Best Local Similarity 36.2%; Pred. No. 2.5e-20;
Matches 77; Conservative 32; Mismatches 98; Indels 6; Gaps 4;

QY      50 VAGGRYVNNWNGAETACIEVEGTEGNTTTTRADHDNGNNVA--AYPAIFGCHMGAC 107
Db      49 VIGGRVYVNNMNGGTSATQCV-TATDSG-FRYTQAGSVPTNGAPKSYSVNCGHYTNC 106
QY      108 TNSGGLPRVQELSDVTSWLTPTTGRNNAAYDIWSPVTSNGSGGAEIMILMNM 167
Db      107 SPETALPARISGISAPSSISYGFVDNAYNASYDITLDPPTPTDS--VNRTIIMTFMR 164
QY      168 NGGVGSGRVAIVELAGATWVWYADMDNYIAYRRTPPTSVELDKAFIDAVANG 227
Db      165 VGOIQIGSQVGTASVAGRTWEVSGGKNTINDLSVAPASMSNMFVMDVRAIVANG 224
QY      228 YIRPEWYLAHETGFELEGGAGLSADFSVTYV 260
Db      225 LAGNDWYLTSTIQAGFEPMONGAGLAVNSFSSTYV 257

RESULT 4
ID      008468      PRELIMINARY;      PRT;      377 AA.
AC      008468:
DT      01-JUL-1997 (TREMBLrel. 04, Created)
DT      01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      CEL2 (EC 3.2.1.4).
GN      CELA2.
OS      Streptomyces halstedii.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX      NCBI_Taxid=1944;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=JM8;
RX      MEDLINE=97307849; PubMed=9182697;
RX      Garda-Salas A.L., Fernandez-Abalos J.M., Sanchez P., Ruiz-Arribas A.,
RA      Santamaria-Sanchez R.I.;
RT      "Two genes encoding an endoglucanase and a cellulose-binding protein
RT      are clustered and co-regulated by a TTA codon in Streptomyces
RT      halstedii JM8.";
RL      Biochem. J. 324:403-411(1997).
DR      EMBL: U51222; AAC45429.1; -.

```



```

QY 171 VMBGSRVAVELAGATWEVYADWDMNNTAYRRTPPTTSVSELDLKAFLDDAVARCYIR 230
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 171 IQPIGSPVGNASVGGRTWVEWWSGGNSNDYLSFVAPSAISGSPVDVDFRATVARGLAE 230
QY 231 PEWYIHAVETGFELMEGGAGLRSDAFSVTYO 261
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 231 NDWLTISVQAGFEPMQNGAGLAVNSFSSTVE 261

RESULT 7
Q9X602 PRELIMINARY; PRT; 384 AA.
ID O9X602;
AC O9X602;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CELLULASE.
GN CEL51.
OS Streptomyces viridosporus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=67581;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T7A;
RA Ramchandran S., Crawford D.L.;
RT "Characterization and sequence analysis of two genes involved in
RT cellulase degradation in Streptomyces viridosporus T7A, and its
RT expression in Escherichia coli.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF130408; AND25090.1; -.
DR HSSP: P07986; IEXG.
DR InterPro: IPR001919; CBD_2.
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF00553; CBD_2; 1.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
SQ SEQUENCE 384 AA; 40918 MW; D3968B6EBDEDE65 CRC64;

Query Match 23.4%; Score 338.5; DB 2; Length 384;
Best Local Similarity 32.1%; Pred. No. 3e-19;
Matches 85; Conservative 35; Mismatches 106; Indels 39; Gaps 8;

QY 31 PEPEP-----EPTVELGGRDARDVA-----GGRYRYNNWGAETA 67
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 7 PDPAFCVAVSPSPSPSSRRRRRAQADTTLCPEYGTITIGRYYVQNNRSGSSSP 66
QY 68 QCIEVGLFETGNFTITRADHDNGNNA--AVPATYFGCHMGACTSNGSLPRVQELSDVTR 125
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 67 QCV-TATDTG-FRLTQADGSVPTNGAPKSTPSVFNCHITNCSPGTRKLPARISGISAPL 124
QY 126 -----SWTLPTITGR----WNAAVDWFSPYTSNGSGYSGAELIMLNNGVMPG 175
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 125 RIXGISAPSSISYGVGAVYNAAYDILMDPTPTDG--VNRETIMFNKKVPIQPIG 182
QY 176 SRVATVELGATWEVYADWDMNNTAYRRTPPTTSVSELDLKAFLDDAVARCYIRPEWYL 235
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 183 SOVGATVVGGRFQWVWSGGNSNDYLSFVAPSAIESWSEVDVDFRETVARQMAQDWYL 242
QY 236 HAVETGFELMEGGAGLRSDAFSVTY 260
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 243 TSVQAGFEPMQNGAGLAVNSFSSTV 267

RESULT 8
Q31030 PRELIMINARY; PRT; 264 AA.
AC Q31030;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

```

```

DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.
GN CELB.
OS Pectobacterium carotovorum subsp. carotovorum.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LY34;
RX MEDLINE=98096373; PubMed=9434760;
RA Park Y.W., Lim S.T., Cho S.J., Yun H.D.;
RT "Characterization of Erwinia carotovora subsp. carotovora L134 endo-
RT 1,4-beta-glucanase genes and rapid identification of their gene
RT products.";
RL Biochem. Biophys. Res. Commun. 241:636-641(1997).
DR EMBL: AF025769; AAC02965.1; -.
DR InterPro: IPR002594; Glyco_hydro_12.
DR Pfam: PF01670; Glyco_hydro_12; 1.
DR ProDom: PD004316; Glyco_hydro_12; 1.
KW Signal; Hydrolase.
FT SIGNAL 1 36
FT CHAIN 37 264
FT STRAIN-A3(2);
SQ SEQUENCE 264 AA; 29634 MW; 70EA366B8443CEB CRC64;

Query Match 16.8%; Score 242.5; DB 2; Length 264;
Best Local Similarity 29.8%; Pred. No. 8.6e-12;
Matches 68; Conservative 39; Mismatches 90; Indels 31; Gaps 9;

QY 54 RRYVNNWGAETVACQIEVGLFETGNFTITRAD-----HDGNNVAAVPAITFGCHGA 106
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 48 KYVFNNWGRKDEVK---GMOQTFEYNSPTSGMMHWPSSSVKAYPSLISGWHMTA 103
QY 107 C-TSNGSLPRVQELSDVRTSWTLPTTGRWNAAYDIWFSPYTSNGSGYSGAELIMWL 165
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 104 YTEENSGLPKILSSNKSITSNTVYLSIKSTATLYAAVDVHFHTDKASMSSTPDELIMWL 163
QY 166 NMNGGVMPGSGRVAIVELAGATWEVYADW-----DMNTAYRRTPPTTSVSELDLKA 218
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 164 N-NTNAGPAGDYETVFLDGSSMNV-FKGMINAGNGGNNVFSVRTSNTNSAS-LNIRH 220
QY 219 FIDDAVARGYIRPEW-----YLHAVETGFELMEGGAGLRSDAFSVTYO 261
      :| | | | | :| | | | | :| | | | | :| | | | |
Db 221 FTNDVQ-----KKWMSNAKYISSVELGTEIRGGDQIDITKMSDVK 264

```

```

RESULT 9
Q9KYS4 PRELIMINARY; PRT; 244 AA.
ID Q9KYS4;
AC Q9KYS4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE SECRETED SUGAR HYDROLASE.
GN SC5H4.15.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Murphy L., Harris D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;

```

RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinasht H., Hopwood D.A.:  
 RA "A set of ordered cosmids and a detailed genetic and physical map for  
 RA the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL355913; CAB91127.1.  
 DR InterPro: IPR002594; Glyco\_hydro\_12.  
 DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
 DR ProDom: PD004316; Glyco\_hydro\_12; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 244 AA; 26123 MW; 2CA63242798EAD4 CRC64;

Query Match 14.1%; Score 204.5; DB 2; Length 244;  
 Best Local Similarity 26.7%; Pred. No. 8.5e-09;  
 Matches 60; Conservative 52; Mismatches 78; Indels 35; Gaps 11;

OY 42 CGRMAROVAGRRVYVNNWGAET-AOCI--EVLGTGNTFTTRADHDGNNVAAYPAI 98  
 DB 46 CDQWNTSLNG--YTLVNNWGSAGSGOCVANSSTDWGVW---ADHPNTGIKSYF-- 97  
 OY 99 YFGCHMGACTSNGLPRVOELSDVTSWTLPTTGRNNAAYDIWFSPTNSGNGSGG 158  
 DB 98 -----NAKKVINKPITSLSSLTSSVNTVPSSGAANTSYDID-----DTDYD--- 140  
 OY 159 AELMTLWNGGVMPGSGRVATVELAGATWEYVADMDNNTY-AYRRTPTTS--VSELD 215  
 DB 141 -EIMLMVNHGAVGVLGTFQSGVGLGHTWDYKGNNGANEFSEFLRNSDSNGTVNLP 199  
 OY 216 LKAFLDDAVARGYIRPEWYLHAVETGTFELMEGAGLRSADEFSTV 260  
 DB 200 ILKWLKD--TKGMGNE-TIGDVPFGYEITSSAGGLDFNTNLTIV 241

RESULT 10  
 ID 060033 PRELIMINARY: PRT: 274 AA.  
 AC 060033.

DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE ENDO-1,4-BETA-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULOSE)  
 DE (ENDOGLUCANASE) (CARBOXYMETHYL CELLULOSE).  
 GN CELB OR TM1525.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 (DSM3109);  
 RX MEDLINE=96425879; PubMed=8828221;  
 RA Liebl W., Rulle P., Bronnenmeier K., Riedel K., Lotzpeich F.,  
 RA Greif I.;  
 RT "Analysis of a Thermotoga maritima DNA fragment encoding two similar  
 RT thermostable cellulases, CelA and CelB, and characterization of the  
 RT recombinant enzymes";  
 RL Microbiology 142:2532-2542(1996).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Lither K.D., Garrett M.M.,  
 RA Stead A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of Thermotoga maritima";  
 RL Nature 399:323-329(1999).  
 DR EMBL: Z69341; CAA93274.1; -.  
 DR EMBL: AE001800; AAD36592.1; -.  
 DR TIGR: TM1525; -.

DR InterPro: IPR002594; Glyco\_hydro\_12.  
 DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
 DR ProDom: PD004316; Glyco\_hydro\_12; 1.  
 KW Signal; Hydrolase; Glycosidase; Complete proteome.  
 FT SIGNAL 1 17  
 FT CHAIN 18 274  
 SQ SEQUENCE 274 AA; 31734 MW; 72C0A5ACAD7BA358 CRC64;

Query Match 14.0%; Score 202; DB 16; Length 274;  
 Best Local Similarity 28.3%; Pred. No. 1.5e-08;  
 Matches 73; Conservative 40; Mismatches 87; Indels 58; Gaps 16;

OY 38 TVEIGRMDARDVAGRRVYVNNWGAETACIEVLEGTGNTFTTRADHDGNNVAAYPA 97  
 DB 38 TWEL-NFWNVKSYEG-----ETWLKFDGEKVEFYADLYNVLQNP-----SVHGYPE 85  
 OY 98 IYFGCH-WGACTSNNG--LPRVOELSD-VRTSWTL-----PTITGRNNAAYDIWFS 146  
 DB 86 IYGYKPMAG--HNSGVEFLPVKRDLPDYVYLDYSIWYENNLPI-----NLAEWTIT 138  
 OY 147 PVTNSGNGYSGGAEELIMLMNGGVMPGSGRV---ATVELAG---ATWEYVADMDN 198  
 DB 139 RSPDQTSVSSGDALIMWF-YNNVLMFGGQKDEFTTYEINGVAKQETKMDYFAPMGMD 197  
 OY 199 YIAYRRTPTTSVS-ELDIKAFIDDA---VARGYIRPEWYLHAVETGTFELMEGA----- 249  
 DB 198 YLAFRLTTPMKRGKKIKVKKDFQKRAAEVKKHSTRID---NFEELYCVMWIGTEFGDP 254  
 OY 250 -----GLRSADFSVTY 260  
 DB 255 NTTAKFGWTFRDFSEYEV 272

RESULT 11  
 ID P96492 PRELIMINARY: PRT: 274 AA.  
 AC P96492.

DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE ENDO-1,4-BETA-GLUCANASE B (EC 3.2.1.4) (CELLULOSE)  
 DE (CARBOXYMETHYL CELLULOSE).  
 GN CELB.  
 OS Thermotoga neapolitana.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID=2337;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC BOX J.D., Eveleigh D.E.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-137 FROM N.A.  
 RC STRAIN=Z2706-MC24;  
 RA Zverlov V.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-137 FROM N.A.  
 RC STRAIN=Z2706-MC24;  
 RX MEDLINE=93356813; PubMed=8352795;  
 RA Dakhova O., Kurepina N., Zverlov V., Svetlichny I V.,  
 RA Velikodvorskaya G.;  
 RT "Cloning and expression in Escherichia coli of Thermotoga neapolitana  
 RT genes coding for enzymes of carbohydrate substrate degradation.";  
 RL Biochem. Biophys. Res. Commun. 194:1359-1364(1993).  
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
 LINKAGES IN CELLULOSE.

CC EMBL: U93354; AAC95060.1; -.  
 DR EMBL: Z86103; CAB06782.1; -.  
 DR InterPro: IPR002594; Glyco\_hydro\_12.  
 DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
 DR ProDom: PD004316; Glyco\_hydro\_12; 1.  
 KW Hydrolase; Glycosidase.





01-DEC-2001 (Tremblrel. 19, last annotation update)  
DE ENDO-BETA-1,4-GLUCANASE (EC 3.2.1.4).  
GN EGL.  
OS Trichoderma reesei (Hypocrea jecorina).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreales; Hypocreaceae; Hypocrea.  
OX NCBI\_TaxID=51453;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=QM9414;  
RA Okada H., Tada K., Sekiya T., Yokoyama K., Takahashi A., Tonda H.,  
Kumagai H., Morikawa Y.,  
RT Molecular characterization and heterologous expression of the gene  
RT encoding a low-molecular-mass endoglucanase from Trichoderma reesei  
RT QM9414.  
RL Appl. Environ. Microbiol. 64:55-563(1998).  
DR EMBL: AB003694; BAA20140.1;  
DR InterPro: IPR002594; Glyco\_hydro\_12.  
DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
DR Prodom: PD004316; Glyco\_hydro\_12; 1.  
KW Hydrolyase; Glycosidase.  
SQ SEQUENCE 234 AA; 25159 MW; DF476EDE384ADD1 CRC64;

Query Match 12.2%; Score 176.5; DB 3; Length 234;  
Best Local Similarity 28.6%; Pred. No. 1.4e-06;  
Matches 68; Conservative 34; Mismatches 93; Indels 43; Gaps 14;

QY 42 CGRMDARDVAGRGYRVINNVGAETQ---CI-EVGLGTGNFTITRAD---HDNGNNVAA 94  
DB 20 CDQM--ATFTGNGITYVSNMLMGASAGSGFCYTAVSLSG--ASWHADQMWSGGQNNVKS 75  
QY 95 YPAIVEGCHWGACTNSGLP--RRVOELSDVPT--SWTLPTITGRNNAAYDIWFSPVTN 150  
DB 76 YQ-----NSQIAIPQKRTVNSISSMPTTASMSYSG--SNIRANVAYDL-FTANP 122  
QY 151 SGNGYSGGAEIMTLWMNGVMPGSGRVATVELAGATWEVWY----ADWDNNTIATIRRTT 206  
DB 123 NHVTYSGDELMTLKYGIGIPISGSGTGVNNGQSWTLTYGYNGAMQVYSFVAQNTIT 182  
QY 207 PPTSVELDLKAFI---DDAVARGYIRPEWYLHAVETGFELMEGAGLSADPSVTV 260  
DB 183 NYSG----DYKNFPNIRDN---KGYNAAGQVLYSTQFTGTEPFTGSGTLNVAWSIASI 233

## RESULT 15

ID 013454 PRELIMINARY; PRT; 239 AA.  
AC 013454;  
DT 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)  
DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4).  
GN CELA.  
OS Aspergillus oryzae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurobiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5062;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=KBN616;  
RX MEDLINE=97161783; PubMed=9008887;  
RA Kitamoto N., Go M., Shibayama T., Kimura T., Kito Y., Ohmiya K.,  
Tsukagoshi N.,  
RT Molecular cloning, purification and characterization of two endo-1,4-  
RT beta-glucanases from Aspergillus oryzae KBN616.  
RL Appl. Microbiol. Biotechnol. 46:538-544(1996).  
DR EMBL: D83731; BAA22588.1;  
DR InterPro: IPR002594; Glyco\_hydro\_12.  
DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
DR Prodom: PD004316; Glyco\_hydro\_12; 1.  
KW Hydrolyase; Glycosidase.  
SQ SEQUENCE 239 AA; 26096 MW; C0F850E5DEB455D CRC64;

Query Match 11.9%; Score 172.5; DB 3; Length 239;  
Best Local Similarity 27.3%; Pred. No. 2.9e-06;  
Matches 63; Conservative 37; Mismatches 86; Indels 45; Gaps 11;  
QY 40 ELGGRMDARDVAGRGYRVINNVGAET---AOCIEVGLGTGNFTITRADHDNGNNVAA 96  
DB 18 ELCAQYDS--ASSPPYSVNNMLMGQDSGSGCYVY-----DNLSSGA-- 59  
QY 97 AITFGCHW-----GACTSNSGLPRRYOEL--SDVRTSWTLPTITGRNNAAYDI 143  
DB 60 AWHTTWTWNGEGSVKSYNSAVTFPKLIVSDVQSIPDVEWSQDNTNNA---DVAYDL 116  
QY 144 WFSPPVNSGNGYSGAEELMTLWMNGVMPGSGRVATVELAGATWEVWY----ADWDNN 198  
DB 117 -FTAADQNNHTYISGDYELMIMLARVGTIQIDTATVEGHTWELWGTITQAGAEQK 175  
QY 199 YIAYRRTPTPTSVELDLKAFIDDAVAR-GYIRPEWYLHAVETGFELMEGG 248  
DB 176 TYSFVSATPINTFGG-DIKKFPDYITSKHSFPASQYLLINMGFTGTEPFTG 225

Search completed: August 15, 2002, 10:57:39  
Job time: 393 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:51:03 ; Search time 53.28 Seconds

(without alignments)  
462.807 Million cell updates/sec

Title: us-10-003-759-2\_COPY\_40\_261

Perfect score: 1234  
Sequence: 1 ELGGRWDARVDVAGGRYVIN.....FEIMEGAGLRSADEFVTYQ 222

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1106.5	89.7	260	20	AAV06369
2	1106.5	89.7	260	21	AA1882
3	1106.5	89.7	260	21	AA184347
4	470.5	38.1	105	20	AAV06342
5	396.5	32.1	371	20	AAV06367
6	396.5	32.1	371	20	AAV08473
7	396.5	32.1	371	21	AA1880
8	396.5	32.1	371	21	AAV84345
9	396.5	32.1	371	21	AAV67496
10	396.5	32.1	386	21	AAV67497
11	343	27.8	381	20	AAV06368

12	343	27.8	381	21	AA1881	Rhodothermus marin
13	329	26.7	429	21	AAV84346	Amino acid sequenc
14	255.5	20.7	261	20	AAW88462	Bacillus lichenifo
15	252.5	20.5	264	20	AAV06370	Erwinia carotovora
16	252.5	20.5	264	21	AA1883	Amino acid sequenc
17	252.5	20.5	264	21	AAV84348	Aspergillus aculea
18	224.5	18.2	259	20	AAV06351	Aspergillus aculea
19	224.5	18.2	259	21	AA1864	Amino acid sequenc
20	224.5	18.2	259	21	AAV84329	Aspergillus kawach
21	214	17.3	239	20	AAV06353	Aspergillus kawach
22	214	17.3	239	21	AA1866	Amino acid sequenc
23	214	17.3	239	21	AAV84331	Humicola insolens
24	212	17.2	254	21	AA1869	Humicola insolens
25	212	17.2	254	21	AAV84334	Amino acid sequenc
26	210.5	17.1	246	20	AAV06366	Emeticella deserto
27	210.5	17.1	246	21	AA1879	Actinomycete 11A8
28	210.5	17.1	246	21	AAV84344	Amino acid sequenc
29	209	16.9	254	21	AA1868	Humicola grisea EG
30	209	16.9	254	21	AAV84333	Humicola grisea EG
31	209	16.9	254	21	AAV07558	Humicola grisea en
32	209	16.9	254	22	AAU07584	Humicola grisea en
33	208.5	16.9	254	22	AAU07583	Humicola grisea en
34	201.5	16.3	255	20	AAV06356	Humicola insolens
35	199	16.1	104	20	AAV06337	Streptomyces sp. E
36	192.5	15.6	253	20	AAV06355	Humicola grisea EG
37	192.5	15.6	348	20	AAV06362	Gliocladium roseum
38	192.5	15.6	348	21	AA1875	Gliocladium roseum
39	192.5	15.6	348	21	AAV84340	Amino acid sequenc
40	191.5	15.5	234	22	AAU07582	Trichoderma reesei
41	184.5	15.0	244	19	AAW68593	Tiarospora phas
42	184.5	15.0	244	21	AAV44341	T. phaseolina xylo
43	183	14.8	244	20	AAV06359	Fusarium javanicum
44	183	14.8	244	21	AA1872	Fusarium javanicum
45	183	14.8	244	21	AAV84337	Amino acid sequenc

#### ALIGNMENTS

RESULT 1	AAV06369	standard; Protein; 260 AA.
ID	AAV06369	standard; Protein; 260 AA.
XX	AAV06369;	
AC	AAV06369;	
DT	06-SEP-1999 (first entry)	
XX		
DE	Rhodothermus marinus Egit-1-like cellulase.	
KW	Cellulase; endoglucanase; Egit-1; textile; feed additive; baking;	
XX	Food processing; grain wet milling; pulp; paper.	
OS	Rhodothermus marinus.	
XX		
PN	MO9931255-A2.	
XX	24-JUN-1999.	
PD		
XX		
PF	14-DEC-1998; 98MO-US26552.	
XX		
PR	16-DEC-1997; 97US-0991720.	
XX		
PA	(GENV) GENENCOR INT INC.	
PI	Bower BS, Fowler T, Phillips JI;	
XX		
DR	WPI, 1999-395187/33.	
XX		
PT	Egit-1 like cellulase	
XX		
PS	Example; Fig 6; 47pp; English.	
XX		
CC	The present polypeptide represents a full-length sequence of a	

novel EglIII-like cellulase isolated from *Rhodothermus marinus*. It was deduced from a gene sequence isolated from genomic DNA using PCR primers (see AA551980-91) based on conserved motifs (see AA106325-29) of *Trichoderma reesei* EglIII cellulase and related enzymes. PCR has been used to identify novel EglIII-like enzymes, including the present protein, from bacterial and fungal sources (see AA106331-70). The sequence shows homology to T. reesei EglIII (see AA106330). Also provided by the invention are vectors, host cells and methods for the recombinant production of such enzymes, which can be used in the treatment of cellulose-containing textiles, as feed additives, in the treatment of wood pulp, in the reduction of biomass to glucose, in the stone washing of indigo dyed denim, or as laundry detergent components (all claimed).

**SQ Sequence 260 AA;**

Query Match	89.7%	DB: 20;	Length	260;					
Best Local Similarity	88.6%	Pred. No.	9.6e-107;						
Matches	202;	Conservative	4;	Mismatches	9;	Indels	13;	Gaps	2;

```

OY 1 ELGRMDAPRVAGRGYRVIVINWVCAEFAOICEVGLFENGTITIRADHDNCNNVAAAPAY 60
Db 40 elgrvrdardvaagryrvvlnwvgaeeaqcevglevtegnfllcttadhdngmnaaypaly 99
OY 61 FCGHMG-----CTSNGLPRRQELSDVRTSTLTPITTRGMNAAYDIWFSVPSVNSGN 114
Db 100 fchwaparalrldcaaragavrreheld-----vplltgltgmaaydlwfsprvnsng 152
OY 115 GYSGGALMTLWLNWNGVMEGGSFVATVELAGATWEYVWADMDMNYIARRTTPTTSVSE 174
Db 153 gysggaeltmwlwnvgvmpgsrvalvelagatwewyadwmyayrrtpttcvase 212
OY 175 LDKAFITDDAVARGYIRPEWYLAHVEGFLWEGGAGLSADSPSYNQ 222
Db 213 ldkafitddavargyirpewylhavelgfelewggagllradsfsyvaq 260

```

## RESULT 2

ID AAB14882 standard; Protein; 260 AA.

AC AAB14882

DT 21-NOV-2000 (first entry)

DE Emericella desertoru EGI1-like cellulase.

Emerlicella desertoru; Trichoderma reesel; endoglucanase III; EGIII;

KW wood pulp treatment; feed additive; detergent.

OS Emerlcella desertoru.  
xy

PN WO2000037614-A2.

PD 29-JUN-2000.

PF 12-NOV-1999; 99WO-US26704.  
yy

PR 18-DEC-1998; 98US-0216295.  
XX

PA (GEMV ) GENENCOR INT INC.

PI Mitchinson C, Wendt DJ;

DR WPI; 2000-482483/42.

Novel endoglycanase

PT specified positions in the wild form of endoglucanase III  
XX  
PS Example 1; Fig 3; 52pp; English.

XX The present sequence is a cellulase related to endoglucanase III (EgIII) from *Trichoderma reesei*. EgIII-like genes were isolated from genomic DNA libraries constructed from various microorganisms by PCR. The isolated clones showed significant homology to EgIII from *T. reesei*. Certain CC substitution and deletion mutations have been incorporated into EgIII and CC EgIII-like cellulases to produce variant enzymes with improved stability CC e.g. increased resistance to temperature stress. The mutants may be used CC in textile and wood pulp treatment, as a feed additive, and for reducing CC biomass to glucose. They are also useful for stonewashing or indigo dyed CC denim and as an agent in laundry and dish detergents.

**SQ Sequence 260 AA;**

Query Match	89.7%	Score 1106.5;	DB 21;	Length 260;
Best Local Similarity	88.6%;	Pred. No. 9.6e-107;		
Matches 202; Conservative	4;	Mismatches 9;	Indels 13;	Gaps 2

QY	1	ELGCRMAPADVAGGRYRIRINNWCAELFAOCCIEVLEENFNITPADHDNGNNVAAYAIY	60
Db	40	elgcrwadardavaggrivrinmwgaecaqclevjlecnfilitradhdngnnvaayaiy	99
QY	61	FGCHMGA-----CTSNGLEPRRVOELSDVRTSMTLPIITGGNNAAYDIWFSPVTSNGN	114
Db	100	fghcmwaparalrdcaaragavrraheld-----vcpitgrnnaaydlwfspvtnsgn	155
QY	115	GYSGGAEMLMTLNNNGVMPGGSRYATVELAGATWEYWIADMDMYNTAYRRTPPTTSYSE	174
Db	153	gysggaeltmltlnnngvmpggsryatvelagatweywidmdynyayrrtpcpttsase	212
QY	175	LDLKAFLDDAVARGYIRPEWTLIAHVEGFLMEGCGALRSADSVPTWO	222
Db	213	ldlkaflddavaryirpewtliahvegfelmegcgalrsadsvptvq	260

### RESULT 3

ID AAY84347 standard; Protein; 260 AA.

AC AAY84347;

DT 12-JUL-2000 (first entry)

DE Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.

KW Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability,

indigo dyed denim; cellulose containing fabric; fabric smoothness;

KW animal feed; wood pulp; paper; grain; biomass reduction; glucose.

OS Rhodothermus marinus.  
VY

PN WO2000014208-A1.

PD 16-MAR-2000.  
yy

PF 24-AUG-1999;  
VY

PR 03-SEP-1998;  
yy

PA (GEMV ) GENENCC

PI Fowler T;

DR WPI; 2000-271052/23.

PT Novel var

XX reducing biomass to glucose --

PS Disclosure; Page 66-67; 73pp; English.  
 XX  
 CC The present sequence represents an endoglucanase III (EgIII)-like  
 CC cellulase. The cellulase has homology to the Trichoderma reesei EgIII  
 CC protein. The variant cellulases have improved temperature stability,  
 CC and improved surfactant stability. The variant cellulases and  
 CC compositions containing them are used in textile processing or cleaning,  
 CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel  
 CC or appearance of cellulose containing fabrics (e.g. improving fabric  
 CC smoothness or removing pills and fibrils). The compositions may also be  
 CC used for the removal of immature or dead cotton from cellulosic fibres  
 CC or fabric, which can cause uneven drying. The cellulase may also be used  
 CC in a detergent composition for washing laundry and dishes and in the  
 CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.  
 CC The enzymes may also be used in the reduction of biomass to glucose.  
 CC  
 XX  
 SQ Sequence 260 AA:

Query Match 89.7%; Score 1106.5; DB 21; Length 260;  
 Best Local Similarity 88.6%; Pred. No. 9.6e-107;  
 Matches 202; Conservative 4; Mismatches 9; Indels 13; Gaps 2;

QY 1 ELGGRWADAVAGGRYVNNWGAETACIEVGLTGNFTIRADHDGNNVAAYPAIY 60  
 DB 40 ELGGRWADAVAGGRYVNNWGAETACIEVGLTGNFTIRADHDGNNVAAYPAIY 99  
 QY 61 FCGHMG-----CTNSGLPRRVOELSDVFTSWLTPITGRRNNAAYDIWFSPTNSGN 114  
 DB 100 fgcHWAPARIdcaaragavrraheld-----vtpiltgrnaaydiwfsptnsn 152  
 QY 115 GYSGAGELMTLWNNGVMEGSGRATVELAGATWEWYADMDNNTAYRRTPTTSYSE 174  
 DB 153 gysggaelmwlnwngvmpgsrvalagatwewyadmdwnyayrrtpttsvse 212  
 QY 175 LDKAFIDDAVARGYIRPEWTLHAVETGFELMEGAGLRSDFSVTVO 222  
 DB 213 ldkafidavargyirpewtlhavegfelwegaglrdsfvtvq 260

RESULT 4  
 AAY06342  
 ID AAY06342 standard; Protein: 105 AA.

AC AAY06342;  
 XX  
 DT 06-SEP-1999 (first entry)  
 XX  
 DE Rhodothermus marinus EgIII-like cellulase (partial sequence).  
 XX  
 KW Cellulase; endoglucanase; EgIII; textile; feed additive; baking;  
 KW food processing; grain wet milling; pulp; paper.  
 XX  
 OS Rhodothermus marinus.  
 XX  
 PN WO9931255-A2.  
 XX  
 PD 24-JUN-1999.  
 XX  
 PF 14-DEC-1998; 98WO-US26552.  
 XX  
 PR 16-DEC-1997; 97US-0991720.  
 XX  
 PA (GEMV ) GENENCOR INT INC.  
 XX  
 PI Bower BS, Fowler T, Phillips JT;  
 XX  
 DR WPI; 1999-395187/33.  
 XX  
 PT EgIII like cellulase  
 XX  
 PS Example; Fig 3; 47pp; English.  
 XX

CC The present polypeptide represents a partial sequence of a novel  
 CC EgIII-like cellulase of Rhodothermus marinus. It was deduced from  
 CC a partial gene sequence isolated from genomic DNA using PCR  
 CC primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)  
 CC of Trichoderma reesei EgIII cellulase and related enzymes. PCR  
 CC has been used to identify novel EgIII-like enzymes, including the  
 CC present polypeptide, from bacterial and fungal sources (see  
 CC AAY06331-70). Also provided by the invention are vectors, host  
 CC cells and methods for the recombinant production of such enzymes,  
 CC which can be used in the treatment of cellulose-containing textiles,  
 CC as feed additives, in the treatment of wood pulp, in the reduction  
 CC of biomass to glucose, in the treatment of indigo dyed denim,  
 CC or as laundry detergent components (all claimed).  
 CC  
 XX  
 SQ Sequence 105 AA:

Query Match 38.1%; Score 470.5; DB 20; Length 105;  
 Best Local Similarity 77.7%; Pred. No. 3.6e-41;  
 Matches 87; Conservative 3; Mismatches 9; Indels 13; Gaps 2;

QY 20 NNWGAETACIEVGLTGNFTIRADHDGNNVAAYPAIYFCGHMG-----CTNSG 73  
 DB 1 nnwgaetaqcievgletgnftrahdngnnvaaypalyfcghwaparalidcaarag 60  
 QY 74 LPRRVOELSDVFTSWLTPITGRRNNAAYDIWFSPTNSGNGSGAGELMTW 125  
 DB 61 avrraheld-----vtpiltgrnaaydiwfsptnsngysggaelmw 105

## RESULT 5

AAY06367  
 ID AAY06367 standard; Protein: 371 AA.

AC AAY06367;  
 XX  
 DT 06-SEP-1999 (first entry)  
 XX  
 DE Streptomyces sp. EgIII-like cellulase.  
 XX  
 KW Cellulase; endoglucanase; EgIII; textile; feed additive; baking;  
 KW food processing; grain wet milling; pulp; paper.  
 XX  
 OS Streptomyces sp.  
 XX  
 PN WO9931255-A2.  
 XX  
 PD 24-JUN-1999.  
 XX  
 PF 14-DEC-1998; 98WO-US26552.  
 XX  
 PR 16-DEC-1997; 97US-0991720.  
 XX  
 PA (GEMV ) GENENCOR INT INC.  
 XX  
 PI Bower BS, Fowler T, Phillips JT;  
 XX  
 DR WPI; 1999-395187/33.  
 XX  
 PT EgIII like cellulase  
 XX  
 PS Example; Fig 6; 47pp; English.  
 XX  
 CC The present polypeptide represents a full-length sequence of a  
 CC novel EgIII-like cellulase of Streptomyces sp. 11A68. It was  
 CC deduced from a gene sequence isolated from genomic DNA using PCR  
 CC primers (see AAX59180-91) based on conserved motifs (see AAY06325-29)  
 CC of Trichoderma reesei EgIII cellulase and related enzymes. PCR  
 CC has been used to identify novel EgIII-like enzymes, including the  
 CC present protein, from bacterial and fungal sources (see AAY06331-70).  
 CC The sequence shows homology to T. reesei EgIII (see AAY06330). Also  
 CC provided by the invention are vectors, host cells and methods  
 CC for the recombinant production of such enzymes, which can be used



CC biomass to glucose. They are also useful for stonewashing or indigo dyed  
CC denim and as an agent in laundry and dish detergents.

XX Sequence 371 AA;

SO

Query Match 32.1%; Score 396.5; DB 21; Length 371;

Best Local Similarity 39.7%; Pred. No. 1e-32;

Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGMWDADVAGGRYRINNMVGAETACIEVGLGTGN-FITPRADHNGNNA--AYP 57  
Db 34 qldcdygtlttlq-dryvvgnmrwtsatqclnv---tgnfcltqdgsvrpnqapksyp 89  
QY 58 AIYFCGHWGACTSNSGLPRVOELSDVRTSWLTPITTTGRMNAAYDIMEFSPYTSNGNGYS 117  
Db 90 svydgchygncprtltpmrissigsapsvsyrytgnvyaaaydiwldp-tprrtngvn 148  
QY 118 GGAELMIMLNMGVMPGSGRYATVELAGATWEVYADMDMNYIARRTPPTTSVSELDL 177  
Db 149 -rteimlwfnrvpqpigspvgtahvggrswewvtcsgnsgndvisflapsaiswscfv 207  
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFELMEGAGLRSDAFSVTV 221  
Db 208 kdfvqavshglatpdytltsiqagfepwegtqlavnsfssav 251

RESULT 8  
AAV84345  
ID AAV84345 standard; Protein; 371 AA.

AC AAV84345;

DT 12-JUL-2000 (first entry)

DE Amino acid sequence of an endoglucanase III (EGIII)-like cellulase.

XX Endoglucanase III; EGIII; EGIII-like cellulase; surfactant stability;

KM cellulase; textile processing; textile cleaning; stonewashing;

KM indigo dyed denim; cellulose containing fabric; fabric smoothness;

KM pill removal; fibril removal; cotton; cellulosic fibre; dying; detergent;

KM animal feed; wood pulp; paper; grain; biomass reduction; glucose.

XX Actinomyces sp.

OS WO200014208-A1.

PN 16-MAR-2000.

XX 24-AUG-1999; 99WO-US19154.

PF 03-SEP-1998; 98US-0146729.

XX (GENEV ) GENENCOR INT INC.

XX Fowler T;

PI

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Disclosure: Page 64-65; 73pp; English.

WPI: 2000-271052/23.

Novel variant endoglucanase III-like cellulases with improved  
surfactant stability and resistance to temperature stress, useful for  
textile processing or cleaning, treating wood pulp, food and grain, and  
reducing biomass to glucose

The present sequence represents an endoglucanase III (EGIII)-like  
cellulase. The cellulase has homology to the Trichoderma reesei EGIII  
protein. The variant cellulases have improved temperature stability,  
and improved surfactant stability. The variant cellulases and  
compositions containing them are used in textile processing or cleaning,  
e.g. stonewashing of indigo dyed denim, and modifying the texture, feel  
or appearance of cellulose containing fabrics (e.g. improving fabric

CC smoothness or removing pills and fibrils). The compositions may also be  
used for the removal of immature or dead cotton from cellulosic fibres  
or fabric, which can cause uneven dying. The cellulase may also be used  
in a detergent composition for washing laundry and dishes and in the  
treatment of animal feed, wood pulp, paper, non-animal foods and grains.  
The enzymes may also be used in the reduction of biomass to glucose.

XX Sequence 371 AA;

SO

Query Match 32.1%; Score 396.5; DB 21; Length 371;

Best Local Similarity 39.7%; Pred. No. 1e-32;

Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGMWDADVAGGRYRINNMVGAETACIEVGLGTGN-FITPRADHNGNNA--AYP 57  
Db 34 qldcdygtlttlq-dryvvgnmrwtsatqclnv---tgnfcltqdgsvrpnqapksyp 89  
QY 58 AIYFCGHWGACTSNSGLPRVOELSDVRTSWLTPITTTGRMNAAYDIMEFSPYTSNGNGYS 117  
Db 90 svydgchygncprtltpmrissigsapsvsyrytgnvyaaaydiwldp-tprrtngvn 148  
QY 118 GGAELMIMLNMGVMPGSGRYATVELAGATWEVYADMDMNYIARRTPPTTSVSELDL 177  
Db 149 -rteimlwfnrvpqpigspvgtahvggrswewvtcsgnsgndvisflapsaiswscfv 207  
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFELMEGAGLRSDAFSVTV 221  
Db 208 kdfvqavshglatpdytltsiqagfepwegtqlavnsfssav 251

RESULT 9  
AAV67496  
ID AAV67496 standard; protein; 371 AA.

AC AAV67496;

DT 19-MAY-2000 (first entry)

DE Actinomyces cellulase protein sequence.

XX Streptomyces sp.

OS

PN

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

WPI: 2000-224344/19.

DR N-PSDB: AA257029.

24-JUN-1998; 98US-0104308.

18-NOV-1998; 98WO-US24649.

28-MAY-1999; 99US-0321981.

(GENEV ) GENENCOR INT INC.

Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;

Claim 1; Fig 1; 72pp; English.

A novel Actinomyces cellulase and related DNA, useful for detergent  
compositions, treating textiles and paper or pulp







CC cellulase. The cellulase has homology to the *Trichoderma reesei* EgitII  
CC protein. The variant cellulases have improved temperature stability,  
CC and improved surfactant stability. The variant cellulases and  
CC compositions containing them are used in textile processing or cleaning,  
CC e.g. stonewashing of indigo dyed denim, and modifying the texture, feel  
CC or appearance of cellulose containing fabrics (e.g. improving fabric  
CC smoothness or removing pills and fibrils). The compositions may also be  
CC used for the removal of immature or dead cotton from cellulosic fibres  
CC or fabric, which can cause uneven dyeing. The cellulase may also be used  
CC in a detergent composition for washing laundry and dishes and in the  
CC treatment of animal feed, wood pulp, paper, non-animal foods and grains.  
CC The enzymes may also be used in the reduction of biomass to glucose.  
XX  
XX Sequence 429 AA:

Query Match 26.7%; Score 329; DB 21; Length 429;  
Best Local Similarity 35.0%; Pred. No. 1.3e-25;  
Matches 75; Conservative 32; Mismatches 101; Indels 6; Gaps 4;

QY 11 VAGGRRVINNWGAETACIEVLEGTGNTTTRADHNGNNA--AYPAIYFGCHWGAC 68  
DB 100 vfgchytqnmrvsgstapqcv-caldtg-fvrlqadgsaptingakpsvfnghytncl 157  
QY 69 TNSGLPRRVOELSDVRTSWTLPTTGRWNAAYDIMFSPVTSNGYSGAGELMTLNM 128  
DB 158 spgtdlprvldtvsapssisfyvgdgvaynasdyldlptardg--vnqetelmwfnr 215  
QY 129 NGGVMPGSGRVATVELAGATWEVYADMDWNYIAYRRTPTTSVSELDKAFIDDAVARQ 188  
DB 216 vspldqldgspvgasvgrtewwsgngsndvlsfvapsalsgwsfdvmdfratvarg 275  
QY 189 YIREPWLIHAVETGFELMEGAGLRSDPSVTYQ 222  
DB 276 laendwyltsvqagfepwngaglavnsisstve 309

## RESULT 14

AAW88462  
ID AAW88462 standard; Protein; 261 AA.

AAW88462;

10-MAY-1999 (first entry)

Bacillus licheniformis xyloglucanase.

xyloglucanase; detergent.

Bacillus licheniformis.

Key Location/Qualifiers

FT Protein 30..261

FT /note= "mature protein, this region is specifically  
claimed in Claim 19"

WO9902663-A1.

21-JAN-1999.

01-JUL-1998; 98WO-DK00290.

24-OCT-1997; 97DK-0001213.

07-JUL-1997; 97DK-0000822.

(NOVO ) NOVO-NORDISK AS.

Bjornvad ME, Jorgensen PL, Outtrup H, Schuelein M;

WPI: 1999-120866/10.

DR N-PSDB; AAX06949.

PT New enzyme preparation comprising a xyloglucanase with an activity

PT of 50 % at pH 7 - useful for improving the properties of cellulosic  
PT fibres, yarn, (non)woven fabric, and rating hemp, jute, flax and  
XX linen fibres

XX Claim 29; Page 71-72; 87pp; English.

XX This polypeptide is an alkaline xyloglucanase obtained from  
CC Bacillus licheniformis ATCC 14580. The enzyme shows optimal  
CC activity at 60 deg C, and retains 50% of its activity after 20 min  
CC at 70 deg. The optimum pH for activity is 5.54, with 32% relative  
CC activity retained at pH 4.59, and 45% at pH 8.49. The heat is  
CC 16.5/sec on xyloglucan at pH 7.5, Km 1.1 g/l. The ratio of maximum  
CC xyloglucanase activity to maximum activity on CM-cellulose is at  
CC least 5:1. An isolated polynucleotide (see AAX06949) encoding the  
CC xyloglucanase can be utilised in the production of recombinant  
CC enzyme. Xyloglucanase preparations are useful for improving the  
CC properties of cellulosic fibres, yarn, (non)woven fabric, and for  
CC rating hemp, jute, flax and linen fibres (claimed). They can also  
CC be used in a process for machine treatment of fabrics, and in  
CC detergent compositions (claimed). The enzyme preparations exhibit  
CC high xyloglucanase activity at alkaline pH without essentially  
CC attacking cellulose or cellulose derivatives.

XX Sequence 261 AA:

Query Match 20.7%; Score 255.5; DB 20; Length 261;  
Best Local Similarity 30.6%; Pred. No. 2.8e-18;  
Matches 70; Conservative 41; Mismatches 85; Indels 33; Gaps 9;

QY 15 RRVINNNWGAETACIEVLEGTGNTTTRADHD-----KNNNAAYPAIYF 61  
DB 45 kyllfnnvsgadqvgswglti-----yhnsdsdmgwwmnpstltvkaypslvs 94  
QY 62 GCHW-GACTSNSGLPRRVOELSDVRTSWTLPTTGRWNAAYDIMFSPVTSNGYSGGA 120  
DB 95 gwhwtegytagsgfprlisdqknlnkvsysangtynaaydlwhnlnkswsdaptcd 154  
QY 121 ELMWLNWNGVWPGGSRVATVELAGATWEVW--YAD---MDWNYIAYRRTPTTSVSE 174  
DB 155 elmlwln-nlnagpagsyvetlsgshwkykgyldagsgkymvnfsflrantcgs-en 212  
QY 175 LDKAFID-AYARGIYREPWLIHAVETGFELMEGAGLRSDPSVTYQ 222  
DB 213 lndftrlnjadsqkwlstkkyssvefgtvgfqlglnsnwdvtvr 261

## RESULT 15

AAW06370  
ID AAY06370 standard; Protein; 264 AA.

AAW06370;

06-SEP-1999 (first entry)

Erwinia carotovora EgitI-like cellulase.

Cellulase; endoglucanase; EgitI; textile; feed additive; baking;

food processing; grain wet milling; pulp; paper.

Erwinia carotovora.

WO9931255-A2.

24-JUN-1999.

14-DEC-1998; 98WO-US26552.

16-DEC-1997; 97US-0991720.

(GENEV ) GENENCOR INT INC.

PI Bower BS, Fowler T, Phillips JT;





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 15, 2002, 10:52:08 ; Search time 23.32 Seconds  
(without alignments)  
232.525 Million cell updates/sec

Title: US-10-003-759-2\_COPY\_40\_261  
1234  
Sequence: 1 ELICGRMDARDVAGGRYRVIN.....FELMEGAGLRSDPSVTVO 222

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1106.5	89.7	260	4	US-09-216-295-23 Sequence 23, Appl
2	396.5	32.1	312	4	US-09-216-295-21 Sequence 21, Appl
3	396.5	32.1	371	4	US-09-104-308-1 Sequence 1, Appl
4	396.5	32.1	371	4	US-09-321-981-1 Sequence 1, Appl
5	396.5	32.1	386	4	US-09-321-981-5 Sequence 5, Appl
6	343.2	27.8	381	4	US-09-216-295-22 Sequence 22, Appl
7	255.5	20.7	261	4	US-09-110-959A-2 Sequence 2, Appl
8	254	20.6	233	1	US-08-032-848C-12 Sequence 12, Appl
9	246	19.9	263	4	US-09-216-295-24 Sequence 24, Appl
10	224.5	18.2	259	4	US-09-216-295-5 Sequence 5, Appl
11	214	17.3	239	4	US-09-216-295-7 Sequence 7, Appl
12	213	17.3	221	1	US-08-032-848C-13 Sequence 13, Appl
13	212	17.2	253	4	US-09-216-295-10 Sequence 10, Appl
14	210.5	17.1	246	4	US-09-216-295-20 Sequence 20, Appl
15	209	16.9	253	4	US-09-216-295-9 Sequence 9, Appl
16	192.5	15.6	348	4	US-09-216-295-16 Sequence 16, Appl
17	183	14.8	243	4	US-09-216-295-13 Sequence 13, Appl
18	179.5	14.5	232	4	US-09-146-770-1 Sequence 1, Appl
19	176.5	14.3	218	1	US-08-032-848C-10 Sequence 10, Appl
20	176.5	14.3	218	2	US-08-438-870-10 Sequence 10, Appl
21	176.5	14.3	218	2	US-08-169-948B-34 Sequence 34, Appl
22	176.5	14.3	218	2	US-08-448-873-34 Sequence 34, Appl
23	176.5	14.3	218	4	US-08-382-452D-34 Sequence 34, Appl
24	176.5	14.3	218	4	US-09-216-295-1 Sequence 1, Appl
25	176.5	14.3	234	1	US-08-032-848C-9 Sequence 9, Appl
26	176.5	14.3	234	1	US-08-438-870-9 Sequence 9, Appl
27	176.5	14.3	234	4	US-09-146-770-3 Sequence 3, Appl

28	176.5	14.3	234	4	US-09-216-295-3 Sequence 3, Appl
29	172	13.9	239	4	US-09-216-295-15 Sequence 15, Appl
30	167	13.5	237	4	US-09-216-295-19 Sequence 19, Appl
31	166.5	13.5	234	4	US-09-146-770-4 Sequence 4, Appl
32	166.5	13.5	234	4	US-09-216-295-4 Sequence 4, Appl
33	163.5	13.2	239	4	US-09-216-295-6 Sequence 6, Appl
34	163.5	13.2	250	4	US-09-216-295-14 Sequence 14, Appl
35	163	13.2	247	4	US-09-216-295-8 Sequence 8, Appl
36	162.5	13.2	223	3	US-08-913-264-1 Sequence 1, Appl
37	162.5	13.2	239	3	US-08-913-264-1 Sequence 13, Appl
38	161.5	13.1	239	4	US-08-849-751-2 Sequence 2, Appl
39	161.5	13.1	239	4	US-09-478-816-2 Sequence 2, Appl
40	160.5	13.0	319	3	US-09-213-042-1 Sequence 1, Appl
41	159.5	12.9	288	4	US-09-216-295-12 Sequence 12, Appl
42	140	11.3	288	3	US-08-995-280C-2 Sequence 2, Appl
43	140	11.3	288	4	US-09-215-042-2 Sequence 2, Appl
44	138	11.2	194	4	US-09-216-295-11 Sequence 11, Appl
45	106.5	8.6	368	3	US-08-591-685-13 Sequence 13, Appl

## ALIGNMENTS

```
RESULT 1
; Sequence 23, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Rhodothermus marinus
; US-09-216-295-23

Query Match      89.7%; Score 1106.5; DB 4; Length 260;
Best Local Similarity 88.6%; Pred. No. 3e+102;
Matches 202; Conservative 4; Mismatches 9; Indels 13; Gaps 2;

QY 1 ELICGRMDARDVAGGRYRVINNWGAETACIEVLEGTGNTIRADHDGNNVAAYPATY 60
DB 40 ELICGRMDARDVAGGRYRVINNWGAETACIEVLEGTGNTIRADHDGNNVAAYPATY 99
QY 61 FGCHMGA-----CTNSGILPRRVOELSDVTSMTLPITTGNNAAVDIWFSPVNSGN 114
DB 100 FGCHMGA-----CTNSGILPRRVOELSDVTSMTLPITTGNNAAVDIWFSPVNSGN 152
QY 115 GYGGAGELMTLWLNMGVMPGSGSRVATVELAGATWEVYADMDMNTIARTPTTSVSE 174
DB 153 GYGGAGELMTLWLNMGVMPGSGSRVATVELAGATWEVYADMDMNTIARTPTTSVSE 212
QY 175 LDKAFIDDAVANGYIRPEWYLHAVETGPELMGAGLRSDPSVTVO 222
DB 213 LDKAFIDDAVANGYIRPEWYLHAVETGPELMGAGLRSDPSVTVO 260

RESULT 2
; Sequence 21, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EGIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
```

;; CURRENT APPLICATION NUMBER: US/09/216,295  
;; CURRENT FILING DATE: 1998-12-18  
;; NUMBER OF SEQ ID NOS: 41  
;; SOFTWARE: FASTSEQ for Windows Version 3.0  
;; SEQ ID NO 21  
;; LENGTH: 312  
;; TYPE: PR1  
;; ORGANISM: Actinomyces 11AG8  
US-09-216-295-21

Query Match 32.1%; Score 396.5; DB 4; Length 312;  
Best Local Similarity 39.7%; Pred. No. 1.2e-31;  
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGRDARDVAGGRVYINNMGAEIACIEVEGLETGN-FTTRADHDGNVNA--AYP 57  
DB 34 QICDRYGTITIQ-DRYVYONNRNGTSATOCINV---TGNGFELTQADGSPVPTNGAKRSYP 89  
QY 58 AIYFGCHMGACTNSGLPRRVOELSDVRTSWTLPTTGRMAAYDIWSPVTSNGSGYS 117  
DB 90 SYVDGCHYGNCAPRTLLPMRISISGAPSSVSRYTGNNGYNAAYDIWLDP-TPTNGVN 148  
QY 118 GGAELMIWLMNMGVAPGSGRVATVELAGATWEVYADMDMYIAVRRTPPTTSVELDL 177  
DB 149 -RTEIMWFRNRGVPVOPISPGVTAHVGRSMVWTGNSGNDVISFLAPSAISSWSFDV 207  
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFEIEMEGAGLSADFSVTY 221  
DB 208 KDFVDQAVSHGLATPDMYLTLSIQAGFEPWEGGTGLAVNSFSAY 251

RESULT 3  
US-09-104-308-1  
; Sequence 1, Application US/09104308  
; Patent No. 6187577  
; GENERAL INFORMATION:  
; APPLICANT: Jones, Brian E.  
; APPLICANT: Van Der Kleij, Wilhelmus A.H.  
; APPLICANT: Van Solingen, Piet  
; APPLICANT: Weyler, Walter  
; TITLE OF INVENTION: No. 6187577el Cellulase Producing Actinomyces,  
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genencor International, Inc.  
; STREET: 925 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1013  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/104,308  
; FILING DATE: 24-JUN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/974,042  
; FILING DATE: 19-NOV-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stone, Christopher L.  
; REGISTRATION NUMBER: 35,696  
; REFERENCE/DOCKET NUMBER: GC539  
; TELEPHONE: 650-846-7555  
; TELEFAX: 650-845-6504  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 371 amino acids  
; TYPE: amino acid

;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
US-09-104-308-1

Query Match 32.1%; Score 396.5; DB 4; Length 371;  
Best Local Similarity 39.7%; Pred. No. 1.5e-31;  
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGRDARDVAGGRVYINNMGAEIACIEVEGLETGN-FTTRADHDGNVNA--AYP 57  
DB 34 QICDRYGTITIQ-DRYVYONNRNGTSATOCINV---TGNGFELTQADGSPVPTNGAKRSYP 89  
QY 58 AIYFGCHMGACTNSGLPRRVOELSDVRTSWTLPTTGRMAAYDIWSPVTSNGSGYS 117  
DB 90 SYVDGCHYGNCAPRTLLPMRISISGAPSSVSRYTGNNGYNAAYDIWLDP-TPTNGVN 148  
QY 118 GGAELMIWLMNMGVAPGSGRVATVELAGATWEVYADMDMYIAVRRTPPTTSVELDL 177  
DB 149 -RTEIMWFRNRGVPVOPISPGVTAHVGRSMVWTGNSGNDVISFLAPSAISSWSFDV 207  
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFEIEMEGAGLSADFSVTY 221  
DB 208 KDFVDQAVSHGLATPDMYLTLSIQAGFEPWEGGTGLAVNSFSAY 251

RESULT 4  
US-09-321-981-1  
; Sequence 1, Application US/09321981  
; Patent No. 6287839  
; GENERAL INFORMATION:  
; APPLICANT: Genencor International, Inc.  
; TITLE OF INVENTION: No. 6287839el Cellulase Producing Actinomyces,  
; TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same  
; FILE REFERENCE: GC540-2  
; CURRENT APPLICATION NUMBER: US/09/321,981  
; CURRENT FILING DATE: 1999-05-28  
; PRIOR APPLICATION NUMBER: PCT/US99/11971  
; PRIOR FILING DATE: 1999-05-28  
; PRIOR APPLICATION NUMBER: 09/104,308  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 08/974,042  
; PRIOR FILING DATE: 1997-11-19  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 371  
; TYPE: PR1  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Nearest "neighbor" - Streptomyces  
US-09-321-981-1

Query Match 32.1%; Score 396.5; DB 4; Length 371;  
Best Local Similarity 39.7%; Pred. No. 1.5e-31;  
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGRDARDVAGGRVYINNMGAEIACIEVEGLETGN-FTTRADHDGNVNA--AYP 57  
DB 34 QICDRYGTITIQ-DRYVYONNRNGTSATOCINV---TGNGFELTQADGSPVPTNGAKRSYP 89  
QY 58 AIYFGCHMGACTNSGLPRRVOELSDVRTSWTLPTTGRMAAYDIWSPVTSNGSGYS 117  
DB 90 SYVDGCHYGNCAPRTLLPMRISISGAPSSVSRYTGNNGYNAAYDIWLDP-TPTNGVN 148  
QY 118 GGAELMIWLMNMGVAPGSGRVATVELAGATWEVYADMDMYIAVRRTPPTTSVELDL 177  
DB 149 -RTEIMWFRNRGVPVOPISPGVTAHVGRSMVWTGNSGNDVISFLAPSAISSWSFDV 207  
QY 178 KAFIDDAVARGYIRPEWYLHAVETGFEIEMEGAGLSADFSVTY 221

Db 208 KDFVDAVSHGLATPDMYLTISQAGFEPMEGTCGLAVNSFSSAV 251

RESULT 5

US-09-321-981-5

Sequence 5, Application US/09321981

Patent No. 6287839

GENERAL INFORMATION:

APPLICANT: Genencor International, Inc.

TITLE OF INVENTION: No. 6287839e1 Cellulase Producing Actinomycetes,

TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same

FILE REFERENCE: GC540-2

CURRENT APPLICATION NUMBER: US/09/321,981

CURRENT FILING DATE: 1999-05-28

PRIOR APPLICATION NUMBER: PCT/US99/11971

PRIOR FILING DATE: 1999-05-28

PRIOR APPLICATION NUMBER: 09/104,308

PRIOR FILING DATE: 1998-06-24

PRIOR APPLICATION NUMBER: 08/974,042

PRIOR FILING DATE: 1997-11-19

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 5

LENGTH: 386

TYPE: PRF

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: Nearest "neighbor" = Streptomyces

OTHER INFORMATION: thermovibaceus

US-09-321-981-5

Query Match 32.1%; Score 396.5; DB 4; Length 386;  
Best Local Similarity 39.7%; Pred. No. 1.5e-31;  
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

1 ELGGRMARVAGRGVYNNVWGAETACCEVLEGTGN-FTIRADHDNNVA--AYP 57  
Db 49 QICRIGTTTQ-DRYVONNRMTSATQCLNV--TGNGEITQAGSVPTNGAPKSY 104

58 AIYFGCHMGACTSNGLPRRYOELSDVRTSWTLPTTGRNNAAYDIWFSPTNSGNGS 117  
Db 105 SYVDGCHYGACAPRTLLPMLISSIGSAPSSVRYTGNQYNNAYDIWLDPTRTGCVN 163

118 GGAELMTLWNGGVMPGSRVATVELAGATWEVYADMDMNTAYRRTPTTSVSELD 177  
Db 164 -RTEIMTFNRVGPVOPIGSPVGFAGHVGGRSMWEVWTSGNSNDVISELAPSAISSWSFDV 222

178 KAFIDDAVARGYRPEWYLAHVEFGFELMEGAGLSADPSVTY 221  
Db 223 KDFVDAVSHGLATPDMYLTISQAGFEPMEGTCGLAVNSFSSAV 266

RESULT 6

US-09-216-295-22

Sequence 22, Application US/09216295

Patent No. 6268328

GENERAL INFORMATION:

APPLICANT: Mitchinson, Colin

APPLICANT: Mendt, Dan J.

TITLE OF INVENTION: No. 6268328e1 Variant EgiII-Like Cellulase Compositions

FILE REFERENCE: GC555

CURRENT APPLICATION NUMBER: US/09/216,295

CURRENT FILING DATE: 1998-12-18

NUMBER OF SEQ ID NOS: 41

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 22

LENGTH: 381

TYPE: PRF

ORGANISM: Streptomyces lividans CelB

US-09-216-295-22

Query Match 27.8%; Score 343; DB 4; Length 381;  
Best Local Similarity 36.5%; Pred. No. 3.1e-26;  
Matches 77; Conservative 31; Mismatches 97; Indels 6; Gaps 4;

14 GRVYVNNWGAETACCEVLEGTGNFTIRADHDNNVA--AYPAIYGCCHMGACTSN 71  
Db 55 GRVYVNNWGAETACCEVLEGTGNFTIRADHDNNVA--AYPAIYGCCHMGACTSN 112

72 SGLPRRYOELSDVRTSWTLPTTGRNNAAYDIWFSPTNSGNGSGAGELMTLWNGG 131  
Db 113 TDLPRVLDVSAAPSSISYGFVDAVYNNAYDIWLDPTARTDG--VNGTEIMTFNRVGP 170

132 VMEGSRVATVELAGATWEVYADMDMNTAYRRTPTTSVSELDKAFIDDAVARGYR 191  
Db 171 IOPIGSPVGTASVGRTEWESGNGSNDVLEFVAPSAISGWSFDVDFVAFVARGIAE 230

192 PEWYLAHVEFGFELMEGAGLSADPSVTY 222  
Db 231 NDWYLTISQAGFEPMEGTCGLAVNSFSSAV 261

RESULT 7

US-09-110-959A-2

Sequence 2, Application US/09110959A

Patent No. 6268197

GENERAL INFORMATION:

APPLICANT: Schuelein, Martin

APPLICANT: Outtrup, Heile

APPLICANT: Bjornvad, Mads Eskelund

TITLE OF INVENTION: Alkaline xylolucanase

FILE REFERENCE: 5206.200-US

CURRENT APPLICATION NUMBER: US/09/110,959A

CURRENT FILING DATE: 1998-07-07

PRIOR APPLICATION NUMBER: 0822/97

PRIOR FILING DATE: 1997-07-07

PRIOR APPLICATION NUMBER: 1213/97

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/054,039

PRIOR FILING DATE: 1997-07-28

PRIOR APPLICATION NUMBER: 60/063,694

PRIOR FILING DATE: 1997-10-28

NUMBER OF SEQ ID NOS: 14

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 261

TYPE: PRF

ORGANISM: Bacillus licheniformis ATCC 14580

US-09-110-959A-2

Query Match 20.7%; Score 255.5; DB 4; Length 261;  
Best Local Similarity 30.6%; Pred. No. 9.1e-18;  
Matches 70; Conservative 41; Mismatches 85; Indels 33; Gaps 9;

15 RYRYVNNWGAETACCEVLEGTGNFTIRADHD-----NGANNVAYPAIYF 61  
Db 45 RYRYVNNWGAETACCEVLEGTGNFTIRADHD-----YHNSDDMGVWMPNNTSTVKAPSIYS 94

62 GCHW-GACTSNGLPRRYOELSDVRTSWTLPTTGRNNAAYDIWFSPTNSGNGSGGA 120  
Db 95 GCHWTEGTYTAGSGFPRLSDOKNINTKYSISANGRYNNAAYDIWLTNTKASDSPA 154

121 ELMTLWNGGVMPGSRVATVELAGATWEVY--YAD--WDNNYJAYRRTPTTSVSE 174  
Db 155 EIMTILN-NTWAGPAGSYVERISIGHSWKYKGYIDAGGCKGNVNSFIRTANTOS-AN 212

175 LDLKAFIDD-AVARGYRPEWYLAHVEFGFELMEGAGLSADPSVTY 222  
Db 213 LNRDFTNYLADSKOMLSKTKYSSVEFGTEFGTGQINISNDVYR 261

RESULT 8

```

US-08-032-848C-12
: Sequence 12, Application US/08032848C
: Patent No. 5475101
:
: GENERAL INFORMATION:
:   APPLICANT: Ward, Michael
:   APPLICANT: Clarkson, Kathleen A.
:   APPLICANT: Weis, Geoffrey L.
:   APPLICANT: Larenas, Edward
:   APPLICANT: Lorch, Jeffrey D.
:   TITLE OF INVENTION: Purification and Molecular Cloning of
:   TITLE OF INVENTION: Eg III Cellulase
:   NUMBER OF SEQUENCES: 20
:   CORRESPONDENCE ADDRESS:
:   ADDRESS: Genecor International
:   STREET: 180 Kimball Way
:   CITY: South San Francisco
:   STATE: CA
:   COUNTRY: USA
:   ZIP: 94080
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patent Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/032.848C
:   FILING DATE: MAR 17 1993
:   CLASSIFICATION: 435
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Horn, Margaret A.
:     REGISTRATION NUMBER: 33,401
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: 415 742-7356
:       TELEFAX: 415 742-7217
:   INFORMATION FOR SEQ ID NO: 12:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 233 amino acids
:       TYPE: amino acid
:       STRANDEDNESS: single
:       TOPOLOGY: linear
:     MOLECULE TYPE: protein
:
: US-08-032-848C-12

```

```

; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Erwinia carotovara
; US-09-216-295-24

Query Match          19.9%; Score 246; DB 4; Length 263;
Best Local Similarity 31.6%; Pred. No. 8,1e-17;
Matches 72; Conservative 40; Mismatches 84; Indels 32; Gaps 11;

```

```

Oy      15 RYRYIANNNWAGMSTAO-----CJEVLGETJENFTIRADHONGNNNAYPALIEGCH 64
Db      48 KYLYFNANNWAGDEIKGMQOITFFINSPLSMG---WNHHPSSSTH----SVKATYSGLVSGWH 100
Oy      65 WGA-CTNSGGLPRRYVOELSDVRTSWTLTLPITTGKMAAAYDIWSPYNSGNGYSGGAELM 1233
Db      101 WTAGTTEHSGLPIDLSSKKSITSVWYYSIKATGYNNAYDIWPTTIDKAMWDSPTDELM 160
Oy      124 IWLN-WNCGVMPGSRVATVELAGATEWYADY-----DNWYIAYRRITPTTSVEL 175
Db      161 IWLNDTNAG--PADDYIETVFLGSSWNV-FKGINADNGGNNV-VSFVHTSGTNSAS-L 215
Oy      176 DKAFFIDAV--ARGYIRREWLIHVAENGFEIMEGAGALRSADSPVWQ 222
Db      216 NRRHTDLYVOTKOWMSDEKTISSVEGTEFLPGDGQIDITTEWYADK 263

```

```

RESULT 10
US-09-216-295-5
; Sequence 5, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wendt, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EgIII-Like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Aspergillus aculeatus
US-09-216-295-5

Query Match      18.2%; Score 224.5; DB 4; Length 259;
Best Local Similarity 29.6%; Pred. No. 1,1e-14;
Matches 76; Conservative 36; Mismatches 92; Indels 53; Gaps 12;

Oy    1 ELGCGWDARDAVGAAGRYRINNVWCAET--AOCIVV--GLEFGNFTIRADHDNG--NNVA 54
Db    20 QLCDOY--ATTYGGYTTINNLMKGRDASGSQCTTVNSASSAGTWSIKMKNSSGENSVK 77
      :|::|||::||::||::||::||::||::||::||::||::||::||::||::||::||
Oy    55 AYPAYIFGCHMGACTSNGL--PRVOELSDVRTSWTLTPITTG--RWNAAYDIWFSPVT 110
Db    78 SY-----ANSGLFENKKLVQSISQIPTTARMSVDNTGIRADVAYHDLFTADI 124
      :|::|||::||::||::||::||::||::||::||::||::||::||::||::||::||
Oy   111 NSGCNYSGCAELMTVLNNNGVMPCGGSKRAYVELAGATWEVWY--ADRMWNITAIRRTPT 169
Db   125 NHWI-WSGDYELMIWLARIYGGVOPIGSQIAATAVDGOTWELMYGANGSOKTYSFYAPPI 183
      :|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Oy   170 TSYSELDELKAFID-----DAYVARGYR-----PEWTLHAVENTGE 204
Db   184 TSF-QGDVNDFPKYLTQNHGFPASSQYLTIQFGETPPTGCPALTSVNMWSAQACFE 242

```





```

Db      92  KSYGYVQKQIQRG-----RISIDINSMTSVSMTYDR-TDIRANVAVDYETIARDPD 144
QY      112  SGNGYSGAELIMILNMNNGSPGGSVAVATVELAGAAVWYADMDNNTLAYRRTPTTS 171
Db      142  HPN-WGDDYELMILARKYGGIYIPGTHSOVNLAGRWDLH-TGYNGMKAYISLIPFSGD 199
QY      172  VSEL--DKAFIDDAVA-RGYIREPWYLHVAETGFEELMEGG-AGLRSADE 217
Db      200  IRRDSCDKIDFENFLERNHGGIPAREQNLIVYQVTECFEGSPAFFTCRDF 249

```

```

RESULT 14
US-09-216-295-20
: Sequence 20, Application US/09216295.
: Patent No. 6268328
:
GENERAL INFORMATION:
: APPLICANT: Mitchinson, Collin
: APPLICANT: Wendt, Dan J.
: TITLE OF INVENTION: No. 6268328e1 Variant EgIII-Like Cellulase Compositions
: FILE REFERENCE: GC555
: CURRENT APPLICATION NUMBER: US/09/216.295
: CURRENT FILING DATE: 1998-12-18
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 20
: LENGTH: 246
:
TYPE: PR1
: ORGANISM: Emeritocella desertoru
: US-09-216-295-20

```

[illegible]

```

US-09-216-295-9

Query Match      16.9%; Score 209; DB 4; Length 253;
Best Local Similarity 28.5%; Pred. No. 3.ee-13;
Matches 68; Conservative % 31; Mismatches 92; Indels 48; Gaps

QY      1  ELGCRMDARDVAGGRYRYRINNMWGAETAAQCEVEGLENGFNITRADHDNCGNVAAPAIY 60
        |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       37  ELYGIV-----SGMGCEYELNNMLMKRDTA-----TSGMCCTIYLDGNG----- 74

QY      61  FGCHWGAQTSNSGHP-----RRVOELSDVVT--SWTLTPITTGNNAAAY 102
        |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       75  -GIDMNTAMENWGAPADNVKNPYVGKQIQGRKSLSDINSMTSVSWYDR-TDLRAVVAY 132

QY      103  DIWTSPTVNSNGTSGCAELMIWLNMGCVPPGSGRAIYVELACATHEWYADMDNYYIA 162
        |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       133  DVETARPDPDHPN-KGGDVEIMIWILARYGGIPICTFHSQVNLACRTDMLW-TGYNGMRY 190

QY      163  YRRATPTTSVSEL--DLKAFIDDAVA--RGYTRPMEYLHAVETGPELWGG--AGLRSADE 217
        |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db       191  YSELPDSGDIDFSCDIDKDFPNYLERNHGYPARQONLIYOVGTECTTGGAARTCDNF 249

```

```
Search completed: August 15, 2002, 10:52:08
Job time: 172 sec
```

```

RESULT 15
US-09-216-295--9
: Sequence 9, Application US/09216295
: Patent No. 6268328
: GENERAL INFORMATION:
: APPLICANT: Mitchinson, Colin
: APPLICANT: Mendt, Dan J.
: TITLE OF INVENTION: No. 6268328el Variant Egitr-Like Cellulase Compositions
: FILE REFERENCE: GC555
: CURRENT APPLICATION NUMBER: US/09/216,295
: CURRENT FILING DATE: 1998-12-18
: NUMBER OF SEQ ID NOS: 41
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 9
: LENGTH: 253
: TYPE: PRT
: ORGANISM: Humicola griseol

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:51:39 ; Search time 29.16 Seconds

(without alignments)  
731.544 Million cell updates/sec

Title: US-10-003-759-2\_COPY\_40\_261

Perfect score: 1234  
Sequence: 1 ELGCRMDARDVAGGRYRIN.....FELMEGAGLRSDPSTVQ 222

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	332	28.5	382	2 JC2571	cellulase (EC 3.2.
2	252.5	20.5	264	2 JU0328	cellulase (EC 3.2.
3	213	17.3	237	2 S12610	cellulase (EC 3.2.
4	200.5	16.2	274	2 A72241	endoglucanase - Th
5	194	15.7	151	2 H70895	hypothetical prote
6	163.5	13.2	239	2 S55831	cellulase (EC 3.2.
7	153	12.4	258	2 H72240	endoglucanase - Th
8	142.5	11.5	332	2 G90291	endoglucanase prec
9	139	11.3	334	2 G90360	endoglucanase prec
10	95.5	7.7	322	2 H90425	hypothetical prote
11	93.5	7.6	673	2 B70528	probable peptidase
12	93	7.5	741	2 G69514	hypothetical prote
13	91.5	7.4	690	2 T27357	hypothetical prote
14	91	7.4	950	2 B87611	ionb-dependent rec
15	88.5	7.2	261	1 S12745	endo-1,4-beta-xyla
16	87	7.1	604	2 B84221	hypothetical prote
17	87	7.1	1449	1 VGTHES	E2 glycoprotein pr
18	85	6.9	957	2 H82261	hemolysin-related
19	84.5	6.8	198	2 AH2306	hypothetical prote
20	84.5	6.8	765	2 S76795	hypothetical prote
21	84	6.8	468	2 AD0585	probable outer mem
22	84	6.8	456	2 H84649	probable DnaJ prot
23	83.5	6.8	475	2 T35697	arabinofuranosidas
24	83	6.7	302	2 H90717	hypothetical prote
25	83	6.7	323	2 H85567	hypothetical prote
26	83	6.7	468	2 H84602	ybfM protein - Esc
27	83	6.7	595	2 T29434	beta-galactosidase
28	82.5	6.7	1090	2 S59077	cellulose 1,4-beta
29	82.5	6.7	1449	1 A43573	E2 glycoprotein pr

30	81.5	6.6	1447	1 VGTHES	E2 glycoprotein pr
31	81.5	6.6	1447	1 VGTHES	E2 glycoprotein pr
32	81	6.6	474	2 G72658	probable Vir B11 A
33	80.5	6.5	318	1 S16016	N-acetylmuramyl-L
34	80.5	6.5	327	2 S34434	antigen 85-B - Myc
35	80.5	6.5	370	2 S75625	hypothetical prote
36	80.5	6.5	520	2 F91080	hypothetical prote
37	80.5	6.5	520	2 G85925	hypothetical prote
38	80	6.5	316	1 A42936	N-acetylmuramyl-L
39	80	6.5	1374	2 AE3259	extracellular seri
40	79.5	6.4	244	2 S23099	lectin I, anti-HO
41	79.5	6.4	453	2 B83380	hypothetical prote
42	79.5	6.4	700	2 F64897	probable ionb-depe
43	79.5	6.4	729	2 T35028	probable glycosyl
44	79.5	6.4	837	1 A56681	penicillin amidase
45	79.5	6.4	986	2 S12021	thermoactive cellu

#### ALIGNMENTS

RESULT 1  
JC2571  
cellulase (EC 3.2.1.4) precursor - Streptomyces rochei (strain A2)  
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase  
C:Species: Streptomyces rochei  
C:Date: 13-Jun-1995 #sequence\_revision 14-Jul-1995 #text\_change 22-Oct-1999  
C:Accession: JC2571; S34392  
R:Perito, B.; Hanhart, E.; Irdani, T.; Iqbal, M.; McCarthy, A.J.; Mastromei, G.  
Gene 148, 119-124, 1994  
A:Title: Characterization and sequence analysis of a Streptomyces rochei A2 endogluca  
A:Reference number: JC2571; M0ID:95011642  
A:Accession: JC2571  
A:Molecule type: DNA  
A:Residues: 1-382 <PER>  
A:Cross-references: EMBL:X73953; NID:q9393391; PIDN:CA52139.1; PID:q9393392  
A:Note: this cellulolytic strain was isolated from the gut of termites  
C:Genetics:  
A:Gene: eg1S  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as  
A:Pathway: cellulose degradation  
C:Superfamily: bacterial cellulose-binding domain homology  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1-37/Domain: signal sequence #status predicted <SIG>  
F:38-382/Product: endoglucanase #status predicted <MAT>  
F:279-380/Domain: bacterial cellulose-binding domain homology <BCB>  
F:280-379/Disulfide bonds: #status predicted

Query Match 28.5%; Score 352; DB 2; Length 382;  
Best Local Similarity 36.2%; Pred. No. 2e-22;  
Matches 77; Conservative 32; Mismatches 98; Indels 6; Gaps 4;

QY	11	VAGGRVYINNNWGAETACIEVLEGTNTTIRADHDGNNVA--AYRAITYGCHHGAC	68
DB	49	VIGGRVYVNNRWGTSATQCV-TATDSG-FRVYQADSVPTNGAPKYPGVFGCHHTNC	106
QY	69	TSNSGLRRVVOELSDVTSWTLTPTTGRNMAADWIFSPVTSNGNGYSGCAELMIWLNW	128
DB	107	SPGALPARISGISASPSISYGFVDNAVYNASDYIDLPTPTDGS--VNRTEIMTFNR	164
QY	129	NGVMPGGSRAVATELAGATWEVYVADMDNVIAYRRTTPTTSVSELDLKAFTDDAVARG	188
DB	165	VGOIPIQISQVGTASVAGRMEVWSGNGTNDVLSFVAPSAMSSMSPDVMDFFRATVARG	224
QY	189	XIREPWYLAHVFGEFELMEGAGLRSDPSTV 221	
DB	225	LAGNDWYLTISIQAFEPWONGAGLAVNSFSTV 257	
RESULT	2		
JU0328			

```

cellulase (EC 3.2.1.4) precursor - Erwinia carotovora subsp. carotovora
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Erwinia carotovora subsp. carotovora
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 08-Oct-1999
C:Accession: J00328
R:Saarilahti, H.T.; Henriksat, B.; Palva, E.T.
Gene 90, 9-14, 1990
A:Title: Cells: a novel endoglucanase identified from Erwinia carotovora subsp. carotovora
A:Reference number: J00328; M0ID:90337352
A:Accession: J00328
A:Molecule type: DNA
A:Residues: 1-264 <SAA>
A:Cross-references: GB:M32399; NID:q148389; PIDN:AAA24817.1; PID:q148390
A:Experimental source: strain SCC3193
A:Note: part of this sequence, including the amino end of the mature protein, was comfirt
C:Genetics:
A:Gene: cells
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-33/Domain: signal sequence #status predicted <Sig>
F:33-264/Product: cellulase #status experimental <Mat>

Query Match      20.5%; Score 252.5; DB 2; Length 264;
Best Local Similarity 31.6%; Pred. No. 3.8e-14;
Matches 72; Conservative 39; Mismatches 86; Indels 31; Gaps 10;

QY 15 RYRVNNWGAETAO-----CIEVGLFETGNTPTTADHDNGNNVAAYATYFGCH 64
DB 48 KLYFNNWVKDKIDKMOQITFYNSPLSMG---WNWHWPSSTH---SVAYPSLVSGWH 100

QY 65 WGA-CTNSGSLPRVVOELSDVRTSWLTPTTGRWNAAYDIMEPSVYNSGNGSGAELM 123
DB 101 WTGAYTENSLELPQLGSSNKSITSNVYISIKATGTYNAAYIWHHTDKAMWDSPTDELM 160

QY 124 TWLN-WNGCWPCCGRVAYVELAGATWEWYADM-----DNVYATRTPTTPTVSSEL 175
DB 161 IWLNDTNAG--PAGDYIEFYFLGDSSWNV-FKGMINADNGGNNVFSFVHTSGTNSAS-L 216

QY 176 DLKAFIDDAV-ARGYIRPEWYLAHVENGFELMEGAGLRSDADFSVTQ 222
DB 217 NHHFTDLYVQTKQMSDEKTVISVEFGEITFGDGOIDITEWRVDVK 264

RESULT 3
SI2610
cellulase (EC 3.2.1.4) precursor - Aspergillus aculeatus
N:Alternate names: endo-1,4-beta-glucanase
C:Species: Aspergillus aculeatus
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C:Accession: SI2610; SI4118; S40186; J00458
R:Ooi, T.; Shimmyo, A.; Okada, H.; Murao, S.; Kawaguchi, T.; Arai, M.
Nucleic Acids Res. 18, 5884, 1990
A:Title: Complete nucleotide sequence of a gene coding for Aspergillus aculeatus cellula
A:Reference number: SI2610; M0ID:91016934
A:Accession: SI2610
A:Molecule type: DNA
A:Residues: 1-237 <COI1>
R:Ooi, T.; Shimmyo, A.; Okada, H.; Hara, S.; Ikenaka, T.; Murao, S.; Arai, M.
Curr. Genet. 18, 217-222, 1990
A:Title: Cloning and sequence analysis of a cDNA for cellulase (FI-CWCase) from Aspergill
A:Reference number: SI4118; M0ID:91064758
A:Accession: SI4118
A:Molecule type: mRNA
A:Residues: 1-237 <COI2>
A:Cross-references: EMBL:X52525; NID:g2287; PIDN:CAA36757.1; PID:g2288
A:Accession: S40186
A:Molecule type: protein
A:Residues: 17-18;42-49, 'X', 51-54, 'X', 66-79;90-111;136-205, 'XX', 208-211 <COI3>
C:Genetics:

```

```

A:introns: 138/2; 212/1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as
A:Pathway: cellulose degradation
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; pyrogutamic acid
F:1-16/Domain: signal sequence #status predicted <Sig>
F:17-237/Product: cellulase #status experimental <Mat>
F:17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experi

Query Match      17.3%; Score 213; DB 2; Length 237;
Best Local Similarity 31.1%; Pred. No. 7.6e-11;
Matches 73; Conservative 40; Mismatches 92; Indels 30; Gaps 12;

QY 1 ELGRMDARDVAGRRYRVINWGAET--AOCTEV--GLETGNFTTTRADHDNG--NNVA 54
DB 20 QLCDQY--ATYTGQVYITNNLNGKADAGSGQCTYVNSASACTSMSTKNNWGGENSVA 77

QY 55 AYPAITFGCHWAGCTNSGL---PRVQELSDVRTSWLTPTTGT--RWNAAYDIMEPSVT 110
DB 78 SY-----ANSGLTFNKKLVISQIPTARMSYDWTGIRADVADLFTAAADI 124

QY 111 NSGNGSGAELMIMLNMGCVMPGSRVATVELAGATWEWY--ADMDNVIYRRTPT 169
DB 125 NHYT-WSGDELMIMLARIGGVQIPISQIATATVDCOTELMIGANGSQKTYSFVAPTP 183

QY 170 TSVSELDLKAIFIDAVA-RCYIRPEWYLAHVENGFELMEGC--AGLRSDADFSVTQ 222
DB 184 TSF-QGDVNDFFKYLQNHGFPPASQYLITLQGTBEFTGTPATLSVMSASAVQ 237

RESULT 4
A72241
endoglucanase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72241
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.E.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; M0ID:99287316
A:Accession: A72241
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-274 <ARN>
A:Cross-references: GB:AE001800; GB:AE000512; NID:g4982090; PIDN:AAD36592.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TMS1525

Query Match      16.2%; Score 200.5; DB 2; Length 274;
Best Local Similarity 27.9%; Pred. No. 1e-09;
Matches 70; Conservative 39; Mismatches 85; Indels 57; Gaps 15;

QY 6 WDARDVAGRRYRVINWGAETAOCTIEVGLFETGNTPTTTRADHDNGNNVAAYATYFGCH- 64
DB 44 WNKSYEG-----ETWLKFDEKEVEFADLVLYLQND-----SMVHGYPEITYYGXP 92

QY 65 WGACTNSG---LPRVQELSD--VRTSWLT-----PTTGTGNNVAAYDIMEPSVTNSGN 114
DB 93 WAG--HNSGVEFLPVKVKDLPDFYVLDDYSIWENNLPi-----NLAETWITRSDQTS 145

QY 115 GYSGAELMIMLNMGCVMPGSRV---ATVLAG-----ATWEVYADMDNVIYRRT 166
DB 146 VSSGDAELIMWF--YNNVLMFGGKVDFTTVEINQKQETKMDVYFAFGMDYLAFLRL 204

QY 167 TPTTSVS-ELDLKAFIDDA---VARGYIRPEWYLAHVENGFELMEGCA----- 210
DB 205 TPKEBEKVKINVDFOKAAEVKKHSTRID---NEELLYFCWELGTGEGDPNTAAKF 261

```

QY 211 GLRSADFSVTY 221  
 Db 262 GWTFRDFSEVY 272

## RESULT 5

hypothetical protein Rv1090 - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C:Accession: H70895  
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Raftery, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A: Reference number: A70500; MID:98295987  
 A: Accession: H70895  
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A: Molecule type: DNA  
 A: Residues: 1-151 <COL>  
 A: Cross-references: GB:AL021897; GB:AL123456; MID:g3256022; PIDN:CAA17206.1; PID:el25196  
 A: Experimental source: strain H37RV  
 C: Genetics:  
 A: Gene: Rv1090

Query Match 15.7%; Score 194; DB 2; Length 151;  
 Best Local Similarity 33.3%; Pred. No. 1.9e-09;  
 Matches 51; Conservative 16; Mismatches 78; Indels 8; Gaps 4;

QY 72 SGLPRVOELSDVTSWTLPTTGRNNAAYDIWF--SPVTSNGXSGAELMIMLWN 129  
 Db 3 TNLPTFGQILSAFTSIDYVPTGVDASDIDLDSTPKTGVN---QOEIMIMFNHQ 58  
 QY 130 GGVMPGSRVATVLAGATWEVYADMDW-NYIAYRRTPTTSSELDKAFIDDAVARG 188  
 Db 59 GSTIOPVSGPVGNNTTECKNFVYMDGNSGMNNAAYVATEP-TEVWSFDVMSFVDHTATME 117  
 QY 189 YIRPEWYLAHVETGFELEMGAGLRSAFVSVTY 221  
 Db 118 PITDSWYLTSLRAGLEPWSDGVLGVDSFAKY 150

## RESULT 6

cellulase (EC 3.2.1.4) precursor - Aspergillus niger  
 N: Alternate names: carboxymethylcellulase I; endo-1,4-beta-glucanase  
 C: Species: Aspergillus niger  
 C: Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 20-Jun-2000  
 C: Accession: S55931; S60657; J50730  
 R: Sakamoto, S.; Tamura, G.; Ito, K.; Ishikawa, T.; Iwano, K.; Nishiya, N.  
 Curr. Genet. 27, 435-439, 1995  
 A: Title: Cloning and sequencing of cellulase cDNA from Aspergillus kawachii and its expression  
 A: Reference number: S55931; MID:96059347  
 A: Accession: S55931  
 A: Molecule type: mRNA  
 A: Residues: 1-239 <SAK>  
 A: Cross-references: EMBL:DJ2901; MID:g217812; PID:g217813  
 A: Note: the source is designated as Aspergillus kawachii  
 A: Accession: S60657  
 A: Molecule type: protein  
 A: Residues: 76-86; 176-186 <SAK2>  
 A: Note: the source is designated as Aspergillus kawachii  
 C: Function:  
 A: Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-  
 A: Pathway: cellulose degradation  
 C: Keywords: blocked amino end; glycosidase; hydrolase; polysaccharide degradation  
 F: 1-16/Domain: signal sequence #status predicted <SIG>  
 F: 17-239/Product: cellulase #status predicted <MAT>

Query Match 13.2%; Score 163.5; DB 2; Length 239;  
 Best Local Similarity 27.3%; Pred. No. 1.2e-06;  
 Matches 63; Conservative 38; Mismatches 97; Indels 33; Gaps 11;

QY 2 LCGRDARDVAGGRYRINNVWG---AETACIEVG--LENGNTTIRADHDNG-NNVAA 55  
 Db 19 MCSQYDS--ASSPPYVSNQNLNMGEGYGTGSCQVYVDKLSSGASGASGASGASGEGTVKS 76  
 QY 56 YPALYFCGHCNACTSNGL---PRVOELSDVTSWTLPTTGRNNAAYDIWFSPVTN 111  
 Db 77 Y-----SNSGLFTFDKLVSDVSSIPTSVTSWSDDTWQADVSTDL-FTANA 122  
 QY 112 SGNGYSGGAEIMLWLNNGVMPGSRVATVLAGATWEVY---ADMDNVIAYRRT 166  
 Db 123 DHATSSGDYELMILWILARYGSVOPIGKQIATATVGKSMEEVWYGSTQAGAEOKTYSFVAG 182  
 QY 167 TPTTSVELDKAFIDDAVA-RGTYREWYLAHVETGFELEMGAGLRSAFVSVTY 216  
 Db 183 SPINSMVG-DIKDFENYLTQNGFPASSQHLITLQCGTEPPTGGPATFTVD 232

## RESULT 7

endoglucanase - Thermotoga maritima (strain MSB8)  
 C: Species: Thermotoga maritima  
 C: Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C: Accession: H72240  
 R: Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Bric-  
 C.M.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
 Nature 399, 323-329, 1999  
 A: Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
 A: Reference number: A72200; MID:99287316  
 A: Accession: H72240  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1-258 <ARN>  
 A: Cross-references: GB:AE001800; GB:AE000512; MID:g4982090; PIDN:AA036591.1; PID:g498  
 A: Experimental source: strain MSB8  
 C: Genetics:  
 A: Gene: TM1524

Query Match 12.4%; Score 153; DB 2; Length 258;  
 Best Local Similarity 27.7%; Pred. No. 1e-05;  
 Matches 51; Conservative 24; Mismatches 87; Indels 22; Gaps 8;

QY 53 VAAVPAIFYGCH-WGACTS-NSGLPRVOELS--DVTSWTLPTTGRNNAAYDIWFSP 108  
 Db 63 VLGIPERYITGKPEWHTAEGSKLPVSVSSKRSVEVSDIHHPSLPLNFAMETWLTR 122  
 QY 109 VTNSNGYSGGAEIMLWLNNGVMPGSR-----VATVELAGATWEVYADMDNMY 160  
 Db 123 EKYGTASIGDVEIMWVEYFN-NLTPGKKEEFTIPVLNGEVECTWELMALAEWMDY 181  
 QY 161 IAYRRTPTTSVS-ELDKAFIDDA-----VANGTYREWYLAHVETGFELEMGAGLR 213  
 Db 182 LAFLKDPVKRKYKFDVRRHFLDAAGRALSSARVKDFEDLYFTWEGTEF--GSPETK 239  
 QY 214 SADF 217  
 Db 240 SAOF 243

## RESULT 8

endoglucanase precursor [imported] - Sulfolobus solfataricus  
 C: Species: Sulfolobus solfataricus  
 C: Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 C: Accession: G90291  
 R: She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch  
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder  
 arett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.





C:Superfamily: prolyl oligopeptidase

Query Match 7.6%; Score 93.5; DB 2; Length 673;

Best Local Similarity 22.6%; Pred. No. 3.3; Matches 56; Conservative 14; Mismatches 61; Indels 117; Gaps 12;

QY 4 GRWDARVAGGRYRVINWGAETACIEVGLTETGNFTTRADHDGNNVAAYPAITFG 62  
DB 467 GRILMAR---GGFYALANIRGGGE-----YG 489  
QY 63 CHNGACSNGLPRRVQELSDVTSWTLPTTGRMAADIMFSPVTSNGNGYSGAEL 122  
DB 490 PGWHTQAMREGGRKQVADFAVAIDLVTRGITTA-----DQLGARSGS-- 532  
QY 123 MIMLNMNGVMPG-----GSRVATVE-----LAGATWEVWYAD---MDW 158  
DB 533 -----NGGLMGIMLTGYEKEFGALVCDVPFLDMKRYHILLACASMAEYGDPEINDDDW 586  
QY 159 NYI-----AYRRTP---TTSVSELDLKAFTIDAVARGYR-----PEW 194  
DB 587 KFISEYSPYONISANRKYPPVLMTTSTR-----DDRYPHGHRKFTALQAAQHPVW 638  
QY 195 YLAHVG 202  
DB 639 YLENIEGG 646

#### RESULT 12

G69514

hypothetical protein AF2119 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Nov-1999

C:Accession: G69514

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
A:; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997

A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kalne, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A:Reference number: A69250; MUID:98049343

A:Accession: G69514

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-741 <KLE>

A:Cross-references: GB:AE000958; GB:AE000782; NID:92689281; PIDN:AA889151.1; PID:9264842

C:Superfamily: Archaeoglobus fulgidus hypothetical protein AF2119

Query Match 7.5%; Score 93; DB 2; Length 741;

Best Local Similarity 24.4%; Pred. No. 4; Matches 49; Conservative 21; Mismatches 51; Indels 80; Gaps 11;

QY 36 ETGNFTTRADHDGNNVAAYPA-----TYFGCHMGACTSNSG-----LPRR-- 77  
DB 204 DTTITLYAAYVNSGNVPPDPVAKFYAEFDSQRIELG--GAWTNSSGIAKLSPINVG 260  
QY 78 -----VOELSDVTSWTLPTTGRMAADIMFSP---VTNSGN-GYSGAELM 123  
DB 261 LSDLRYNFAKIEDVWTCNCAATTTNRAILANVAITGCSYDITLVGRMYSGAD-W 319  
QY 124 IWLNM-----NGVMP-----GSRVATVELAGATWEVWY----- 153  
DB 320 VRNMVYDENSVLYKPIPVQKFTGNR-----ASVIMKYGLDDYCTDPCHREGI 371  
QY 154 -----ADWMNTYAYRRTP 168  
DB 372 YGNFDDADMDGACIAGSTTP 392

#### RESULT 13

T27357

hypothetical protein Y70G10A.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T27357

R:Lloyd, C.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20354

A:Accession: T27357

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-690 <ML>

A:Cross-references: EMBL:AL032660; PIDN:CAA21751.1; GSPDB:GN00021; CESP:Y70G10A.3

A:Experimental source: clone Y70G10A

C:Genetics:

A:Gene: CESP:Y70G10A.3

A:Map position: 3

A:Introns: 61/3; 84/2; 185/1; 250/2; 326/3; 375/1; 398/3; 439/2; 490/3; 628/1; 655/1

Query Match 7.4%; Score 91.5; DB 2; Length 690;

Best Local Similarity 20.5%; Pred. No. 5; Matches 42; Conservative 28; Mismatches 62; Indels 73; Gaps 8;

QY 63 CHNGACTSN-----SGLPVRQV-----LSDVTSWTLPTTGRW 98  
DB 31 CGYGACTPSWLQGFHNKCLLVYGLCAFTQSPVNAIFPVGLSTERRKMTSTHTGIT 90  
QY 99 NAAVD-----IMFSPVTSNG-NGYSGAELMIMLNMNGVMPGSRVATV----- 142  
DB 91 SSWYDFAVLVVFPVYCWGNNGHGR-----WIGMGVIMALSLCALAHNMWDIYHPD 145  
QY 143 -----ELAGATWEVWYADWMNTYAYR-----TTPTSVSELDL 178  
DB 146 VNDLTNQTDFGQCANRDPDPCAGKPHSSMFPYFWMFILQTLHGVSPTLFSIG---T 201  
QY 179 AFIDDAVARGYIREWYLAHVG 203  
DB 202 TYMDENVSQKASPYVLAIRHVLTSF 226

#### RESULT 14

B87611

Toon-dependent receptor [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: B87611

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: B87611

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-950 <STO>

A:Cross-references: GB:AE005673; NID:913424546; PIDN:AAK24886.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC2924

Query Match 7.4%; Score 91; DB 2; Length 950;

Best Local Similarity 22.0%; Pred. No. 7.8; Matches 54; Conservative 32; Mismatches 96; Indels 64; Gaps 14;

QY 2 LCGRWDAVDYAGGRYRVINWGAETACIEVGLTETGNFTTRADH----- 47  
DB 434 LTGEMOYSD-----KLRYDGH-----IGQETSDYDIPISDKRYTAFGLITDYR 478  
QY 48 -DNGNNVAAYPAITFGCHMGACTSNSGLPRVQELSDVTSWTLPTTGRMAADIMF 106  
DB 479 GDGSKNT-----YKMTTIDANN---YRAHET-DFSATYQOTTELKNAEPNAAYDFSD 525



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:58:02 ; Search time 15.93 Seconds

(without alignments)  
539.594 Million cell updates/sec

Title: US-10-003-759-2\_COPY\_40\_261

Perfect score: 1234  
Sequence: 1 ELGCRMDARDVAGGRYVIN.....ELMEGAGLRSDPSVTWQ 222

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	252.5	20.5	264	1	GUNS_ERMCA
2	213	17.3	237	1	GUN_ASPAC
3	163.5	13.2	239	1	GUNA_ASPAK
4	93	7.5	625	1	BGAL_LACSK
5	93	7.5	741	1	VL19_ARCFU
6	88.5	7.2	261	1	XYNA_CLOSA
7	87	7.1	1449	1	VG12_CVPEF
8	86.5	7.0	269	1	EL2_BOVIN
9	83.5	6.8	475	1	ABFB_SMRCO
10	83	6.7	468	1	YBFM_ECOLI
11	82.5	6.7	1090	1	GUXB_CELFI
12	82.5	6.7	1449	1	VG12_CVPMI
13	81.5	6.6	225	1	THTR_CORGL
14	81.5	6.6	1447	1	VG12_CVPRP
15	81.5	6.6	1447	1	VG12_CVPPU
16	80.5	6.5	318	1	ALYS_BRHB3
17	79.5	6.4	700	1	YNCD_ECOLI
18	79.5	6.4	966	1	GUNE_CLOSL
19	79	6.4	475	1	ABFB_STRLI
20	78.5	6.4	233	1	XYNB_MAGRI
21	78.5	6.4	479	1	CB5A_SULSO
22	78.5	6.4	1447	1	VG12_CVPRT
23	77.5	6.3	722	1	ACM1_DROME
24	77.5	6.3	846	1	PAC_ECOLI
25	77.5	6.3	1356	1	HEFI_PODAN
26	76.5	6.2	270	1	KITM_MOUSE
27	76.5	6.2	318	1	ALYS_STRPN
28	76.5	6.2	327	1	A85B_MYCLE
29	76.5	6.2	350	1	MLFI_MALFI
30	76.5	6.2	474	1	SHU1_ECOLI
31	76	6.2	240	1	XYNC_STRLI
32	76	6.2	3396	1	POLG_DENIS
33	75.5	6.1	624	1	TSPE_BSPFV

34	75.5	6.1	641	1	MTIS_SCHPO	074927 schizosach
35	75.5	6.1	966	1	ENV_CAEVC	P31626 caprine art
36	75.5	6.1	3430	1	POLG_MNV	P06935 w genome po
37	75	6.1	441	1	GSA_PROFR	P06774 propionibac
38	75	6.1	945	1	ASPE_MOUSE	P16406 mus musculu
39	74.5	6.0	323	1	A85B_MYCBO	P12942 mycobacteri
40	74.5	6.0	461	1	PSBC_CHLRE	P10898 chlamydomon
41	74.5	6.0	513	1	GUX1_TREIRE	P00725 trichoderma
42	74.5	6.0	581	1	IRR_RAV	P04716 rattus norv
43	74.5	6.0	852	1	POL_BLVJ	P03361 bovine leuk
44	74.5	6.0	872	1	GUXA_CELFI	P04801 cellulomona
45	74.5	6.0	1310	1	VAC3_HELPY	048253 helicobacte

## ALIGNMENTS

RESULT	ID	GUNS_ERMCA	STANDARD	PRT	264 AA.
AC	P16630	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Endoglucanase S precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase S) (cellulase S).				
GN	CELS.				
OS	Erwinia carotovora.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Pectobacterium.				
OX	NCBI_TaxID=554;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN=SCC3193;				
RX	MEDLINE=90337352; PubMed=2379837;				
RA	Saarihahti H.T., Henttinen B., Palva E.T.;				
RT	'celis: a novel endoglucanase identified from Erwinia carotovora				
RT	subsp. carotovora.				
RL	Gene 90:9-14(1990).				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.				
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY H (FAMILY 12 OF GLYCOSYL HYDROLASES).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL: M32399; AAA24817.1; -				
DR	PIR: J00328; J00328.				
DR	InterPro: IPR002594; Glyco_hydro_12.				
DR	Pfam: PF01670; Glyco_hydro_12; 1.				
DR	ProDom: PD004316; Glyco_hydro_12; 1.				
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.				
FT	SIGNAL 1 32				
FT	CHAIN 33 264				
FT	SEQUENCE 264 AA: 29757 MM: E6D61388950C77AA CRC64;				

Query Match 20.5%; Score 252.5; DB 1; Length 264;

Best Local Similarity 31.6%; Pred. No. 3,1e-15;

Matches 72; Conservative 39; Mismatches 86; Indels 31; Gaps 10;

QY	15	RYRYNNWGAETAAQ-----CIEVLEGENFTTPADHDNGNNVAAYPAIYFGCH 64
DB	48	RYRYNNWGAETAAQ-----CIEVLEGENFTTPADHDNGNNVAAYPAIYFGCH 64
QY	65	WGA-CYNSNSGLPRVQGLSVYRTSWTLTPTTGRWNAAYIWPSPVINSNGSGGAEIM 123
DB	65	WGA-CYNSNSGLPRVQGLSVYRTSWTLTPTTGRWNAAYIWPSPVINSNGSGGAEIM 123

ID	NAME	SEQUENCE	START	END	SCORE	ORGANISM	FEATURES
Db	101	WPAGTENSGLPIQSLSNKSI	1	160			
Qy	124	ITLNL-WNGCWPMGGRVAT	1	175			
Db	161	ITLNLITNG--PAGYIEV	1	216			
Qy	176	DKAFITDAV--ARGYIRPE	1	222			
Db	217	NIRHRTDVLVOTKQWMSDE	1	264			
RESULT 2							
AC	GUN_ASPAC	STANDARD:	PRF:	237 AA.			
AC	P22669:						
DT	01-AUG-1991	(Rel. 19, Created)					
DT	01-AUG-1991	(Rel. 19, Last sequence update)					
DT	15-DEC-1998	(Rel. 37, Last annotation update)					
DE	Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)						
DE	(Cellulase) (FI-CMCase).						
OS	Aspergillus aculeatus.						
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;						
OC	Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.						
OX	NCBI_TaxID=5053;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=F-50;						
RC	MEDLINE=91016934; PubMed=2216782;						
RA	Ooi T., Shimmyo A., Okada H., Murao S., Kawaguchi T., Arai M.;						
RT	"Complete nucleotide sequence of a gene coding for Aspergillus						
RT	aculeatus cellulase (FI-CMCase)." ;						
RL	Nucleic Acids Res. 18:5884-5884(1990).						
RN	[2]						
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.						
RC	STRAIN=F-50;						
RC	MEDLINE=91064758; PubMed=2249253;						
RA	Ooi T., Shimmyo A., Okada H., Hara S., Ikenaka T., Murao S.,						
RA	Arai M.;						
RT	"Cloning and sequence analysis of a cDNA for cellulase (FI-CMCase)						
RT	from Aspergillus aculeatus." ;						
RL	Curr. Genet. 18:217-222(1990).						
CC	-1- CATALYTIC ACTIVITY: Endohydrolase of 1,4-beta-D-glucosidic						
CC	linkages in cellulose.						
CC	-1- SUBCELLULAR LOCATION: secreted.						
CC	-1- INDUCTION: BY CELLULOSIC MATERIALS AND HEMICELLULOSES.						
CC	-1- MISCELLANEOUS: WIL ALSO HYDROLYSE 1,4-LINKAGES IN BETA-D-GLUCANS						
CC	ALSO CONTAINING 1,3-LINKAGES.						
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY H (FAMILY 12 OF GLYCOSYL						
CC	HYDROLASIS).						
CC	-----						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -						
CC	the European Bioinformatics Institute. There are no restrictions on its						
CC	use by non-profit institutions as long as its content is in no way						
CC	modified and this statement is not removed. Usage by and for commercial						
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>						
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).						
CC	-----						
CC	EMBL; D00546; BAA00435.1; -						
DR	EMBL; X52525; CAA36757.1; -						
DR	PIR; J00458; J00458.						
DR	PIR; S12610; S12610.						
DR	PIR; S14118; S14118.						
DR	InterPro; IPR002594; Glyco_hydro_12.						
DR	Pfam; PF01670; Glyco_hydro_12.1.						
DR	ProDom; PD004316; Glyco_hydro_12; 1.						
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.						
FT	SIGNAL	1	16	POTENTIAL.			
FT	CHAIN	17	237	ENDOGLUCANASE I.			
FT	MOD_RES	17	17	PYRROLIDONE CARBOXYLIC ACID.			
QO	SEQUENCE	237 AA;	25560 MW;	BE1175571A8A6931 CRC64;			

	Query Match	17.3%	Pred. 213	DB 1	Length 237
	Best Local Similarity	31.1%	Pred. No. 8.2e-12		
	Matches 73	Conservative 40	Mismatches 92	Indels 30	Gaps 12
Qy	1 ELGGRDARDVAGGRYRYINNVYGAET--ADCIYV--GLEGNFTLRADHNG--NNVA 54				
Db	20 QLCDQY--ATYTGYYTTINNMLMGKDAGSGCCTVNSASSAGTSMSTRKMNNGENSESVK 77				
Qy	55 AYPAIYFEGCHMGACNSGL--PRRQELSDYKRSMTLPTTNG--RKNAAVDIMFSYV 110				
Db	78 ST-----ANSGLTFPKKLYVQSISOIPTTARMSYDNTSIRADVAADLTETADI 124				
Qy	111 NSGNGYSGGABELIMVNNNGGYMPCGSRAYETELAGATVEVYV--ADMWNYIAYRRTPY 169				
Db	125 NHVT--WSDGYELMILWLRGQVQIGSOIATFATVYGGOTWELMYGANGSQKTSFYVAPPI 183				
Qy	170 TSVSELDLKAFLIDAVV--RGYIRPEMYLHAVBTGELMEWG--AGRSADFSTQY 222				
Db	184 TSE--QGDVNDFFKYLQNHGEPASSOYLITLQFGTEPPTGCAITLSVMSWSA 237				

[illegible]

DB 19 MCSOYDS--ASSPPYVNONLWGEYQGTSGQCYVDKLSSGASWMTKWTWSGEGTVKS 76  
 QY 56 YPAIFGCHWACTNSGL---PRRVOELSDVTSMTLPTTTG-RWNAAYDIWFSPVTN 111  
 DB 77 Y-----SNSGLFEDKRLKVSIVSSIPSTVWSDODDINQADVSYDL-FTANA 122  
 QY 112 SGNGYSGGAELMTLWNGVMPGSGRVATVELAGATWEVY-----ADMWNVIAYRT 166  
 DB 123 DHATSSGDYELMTLWNLARVGVOPIGKQIAATVATGKGSMEWYGTSTQAGAEQKTYSEVAG 182  
 QY 167 TPTTSELDLKAFLIDAVA-RGIYREPMVLHVENGFELMEGAGARSAD 216  
 DB 183 SPINMSG-DIKDFPNTLTONOGPPASSQHLITLQCGTEPFGPATFTVD 232

## RESULT 4

BGAL\_LACSK STANDARD; PRT; 625 AA.  
 AC 048846;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Beta-galactosidase large subunit (EC 3.2.1.23) (Lactase).  
 GN LACL.  
 OS Lactobacillus sakei.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxID=1599;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DSM 20017;  
 RX MEDLINE=96118231; PubMed=8574399;  
 RA Obst M., Meding E.R., Vogel R.F., Hammes W.P.;  
 RT "Two genes encoding the beta-galactosidase of Lactobacillus sake.";  
 RL Microbiology 141:3059-3066(1995).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-  
 CC galactose residues in beta-D-galactosides.  
 CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC EMBL: X82287; CAA57730.1; -  
 CC DR HSSP: P00722; 1BGU.  
 CC DR InterPro: IPR001649; Glyco\_hydro\_2.  
 CC DR Pfam: PF00703; Glyco\_hydro\_2; 1.  
 CC DR Pfam: PF02836; Glyco\_hydro\_2\_C; 1.  
 CC DR Pfam: PF02837; Glyco\_hydro\_2\_N; 1.  
 CC DR PRINTS: PR001132; GLHYDRASE2.  
 CC DR PROSITE: PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
 CC DR PROSITE: PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
 CC DR HydroLase; Glycosidase.  
 CC FT ACT\_SITE 465 465 PROTON DONOR (BY SIMILARITY).  
 CC FT ACT\_SITE 533 533 NUCLEOPHILE (BY SIMILARITY).  
 CC FT SEQUENCE 625 AA; 72457 MW; C658A3C46136886 CRC64;  
 SO

Query Match 7.5%; Score 93; DB 1; Length 625;  
 Best Local Similarity 17.9%; Pred. No. 0.81;

Matches 44; Conservative 39; Mismatches 87; Indels 76; Gaps 11;

QY 14 GRRRVNNV---WGAETACIEVLETGNTFTTRADHDGNNVAAP---AIYFGCHWG 66  
 DB 342 GKRLVINGVRHMHMBHETRTITAEDEAMDIAQMORNIHNAVRTSHDRLSFYNGC--- 398  
 QY 67 ACSNSGGLPRVQELSDVRSMTLPTTTGRNN--AAVDWFSPVNSG---NGISGGA 120

DB 399 ---DQAGIYMAETNEISHSGWQKGAVEPSWVPGSYDEWEAATIDRARTNEETRNHV 455  
 QY 121 ELMTLWNGVMPGSGRVATVELAGATWEVYADMWNVIAYRTPT----- 169  
 DB 456 SLIFW-----SLGNEYSAGSVLEKMA-----YRKODPTLHYHEGVFA 456  
 QY 170 ---TSVSEL-----DLKAFIDDAVARGYIRPEMVLHAVETG-----F 203  
 DB 497 PEYKATISDVESHMYATPAEIKAYLDNAPQKPLICE-YHHDGNSLIGMQSYIDLLSGY 555  
 QY 204 ELMEGG 209  
 DB 556 DMVGG 561

## RESULT 5

YL19\_ARCFU STANDARD; PRT; 741 AA.  
 AC 028161;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein AF2119 precursor.  
 GN AF2119.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Kleink H.-P., Clayton R.A., Tomb J.F., Hickey E.K., Peterson J.D.,  
 RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Kyrpides N.C.,  
 RA Richardson D.L., Krelavage A.R., Graham D.E., Kyrpides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Ulteck A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton P.M., Spriggs T., Arlatch P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 CC EMBL: AEO00958; AAB89151.1; -  
 CC DR TIGR: AF2119; -  
 CC DR Hypothetical protein; Signal; Complete proteome.  
 CC FT SIGNAL 1 22 POTENTIAL.  
 CC FT CHAIN 23 741 HYPOTHETICAL PROTEIN AF2119.  
 CC FT SEQUENCE 741 AA; 82084 MW; E8C7543552231583 CRC64;  
 SO

Query Match 7.5%; Score 93; DB 1; Length 741;  
 Best Local Similarity 24.4%; Pred. No. 0.98;

Matches 49; Conservative 21; Mismatches 51; Indels 80; Gaps 11;

QY 36 ETGNTFTTRADHDGNNVAAP-----ITFGCHWACTNSG-----LPRR-- 77  
 DB 204 DTDITLTLVAKVNVDPVDPYKFAEFDQSRIPLG---GAWTNSGGLAKLSFIRKNG 260  
 QY 78 -----VQELSDVTSMTLPTTTGRNNAAYDIWFSP---VINSGN-GTSGGAELM 123



```

FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 594 594 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 704 704 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 819 819 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 834 834 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 840 840 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1200 1200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1294 1294 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1311 1311 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1324 1324 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1358 1358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1449 AA; 159957 MW; 971BBAE191FDD1AF CRC64;

```

Query Match Best Local Similarity 7.1%; Score 87; DB 1; Length 1449;

Matches 45; Conservative 20; Mismatches 69; Indels 62; Gaps 10;

```

OY 6 WDARDVAGRRYVNNVGAETACIEVLEGTGNTTTTRADHDGNNV-AAVPAIYFG-- 62
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 88 WD-----YATENSNTWNNHK--QRLNVVNGVYPSITVTTTRNFNSAEGALICICGSP 137
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 63 -----CHNGA-----CTNSGLPRRVOELSDVRSMWLTPTTGRMAAY 102
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 138 PTTTSSLTCCNMGSECRNLNKKPFCPSNS-----EANGCMMLY 176
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 103 DI-WFSPVTSNGVSGGAEIMLW--NNNGVPGGSRVATVELAGATWEVYAD--WD 157
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 177 GLGWFNAVVA---YLHGAGYRISFENMGSTVTLGDMRATILETACTIVDLMWFFNVYD 233
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 158 WNTIATRTPTTSVS 173
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 234 VSYRYVNNKNGTIVVS 249
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 8

EL2\_BOVIN STANDARD; PRT; 269 AA.

```

AC Q29461;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elastase 2 precursor (EC 3.4.21.71).
GN ELA2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Pancreas;
RX MEDLINE=98079203; PubMed=9418008;
RA Gestin M., le Huertou-luron I., Wicker-Planquart C., le Drian G.,
RA Chail J.C., Puigserver A., Guilloteau P.;
RT "Bovine pancreatic preproelastases I and II: comparison of nucleotide
RT and amino acid sequences and tissue specific expression.";
RL Comp. Biochem. Physiol. 118B:181-187(1997).
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Leu-I-Xaa, Met-I-Xaa
CC and Phe-I-Xaa. Hydrolyses elastin.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. ELASTASE SUBFAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: X97635; CAA66231.1; -
DR HSSP: P00766; 1CHG.
DR MEROPS: S01.155; -
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYPSIN_DOM; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 28
FT CHAIN 29 269
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 216 216 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 155 222 BY SIMILARITY.
FT DISULFID 186 202 BY SIMILARITY.
FT DISULFID 212 243 BY SIMILARITY.
SQ SEQUENCE 269 AA; 28856 MW; 8343B97062CF267C CRC64;

```

Query Match Best Local Similarity 7.0%; Score 86.5; DB 1; Length 269;

Matches 57; Conservative 21; Mismatches 63; Indels 93; Gaps 15;

```

OY 4 GRMDARDVAGRRYVNNVGAETACIE-----VG-----LETGNFTTTTRAD--- 46
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 52 GOW--RHTCGG--SLIEQNNVLPFAHCISSSRYRVVVGQSLTVBSGSLTIAVSKSVI 107
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 47 HD-----NGNNVAY-----PAIYFG-----HNGACISNS 72
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 108 HEKWNSQLAOGNDIALKLASSVPLTDKIQGLCPAAGTILPNVYCVYTGGRQDSNG 167
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 73 GLPRVGE-----LSDVPT-----SWTLPTTGRNNAAYDIFSPVTSNGVSGGAEIMI 124
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 168 ALPDIIDQGLIIVDYATYTCNSPMSGTVKTNMICAGD-----GYISSCGDSGGP----- 219
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 125 WLNWNGVMPGSRVATVELAGATWEV-----WYADWNVYIARTPTTSVS 173
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 220 LN-----CQAANROWGVHGVISFGSSIGCNV--YRKSPVTPRVS 256
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 9

ABFB\_STRCO STANDARD; PRT; 475 AA.

```

AC 054161;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-L-arabinofuranosidase precursor (EC 3.2.1.55) (Arabinosidase).
GN ABFB OR SC7H1.02.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D., Parkhill J., Barrrell B.G., Rajandream M.A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

```

CC -1 CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-  
CC arabinofuranoside residues in alpha-L-arabinosides.  
CC -1 PARTWAY: XLYAN DEGRADATION.  
CC -1 SUBCELLULAR LOCATION: Secreted.  
CC -1 SIMILARITY: BELONGS TO FAMILY 62 OF GLYCOSYL HYDROLASES.  
CC -1 SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL: AL021411; CA16189.1; -  
CC InterPro: IPR000772; Ricin\_B\_lectin.  
CC pfam: PF00652; Ricin\_B\_lectin; 1.  
CC SMART; SM00458; Ricin; 1.  
CC DR PROSITE; P550231; Ricin\_B\_Lectin; 1.  
CC Xylan degradation; Hydrolase; Signal; Lactin.  
CC KW SIGNAL 1 37 POTENTIAL.  
CC FT CHAIN 38 475 ALPHA-L-ARABINOFURANOSIDASE.  
CC FT DOMAIN 39 166 RICIN B-TYPE LECTIN.  
CC SQ SEQUENCE 475 AA; 50045 MW; 47E707EE543CA60D CRC64;

Query Match 6.8%; Score 83.5; DB 1; Length 475;  
Best Local Similarity 22.5%; Pred. No. 4.1;  
Matches 43; Conservative 28; Mismatches 77; Indels 43; Gaps 10;

QY 6 WDARDVAGGRTYRINN--VWGAETACIE-VGLETGNFTTRADHDNG----- 50  
DB 110 WSCGGAGNOQMRYNSDGTVVGVESGLCEAAGATNGTAVOLMTCGGGNOKWTLGT 169  
QY 51 ---NNVAAYPAIFGCGHMGACTNSGLPRVQELSDVRSW-FLTPITGRMAAVIDMF 106  
DB 170 PRPDGTICALPSTY---RW-----STGV-----LAQPKSVAALKPFTYTHNGRHLVYG 216  
QY 107 SPVTSNGNGYSGAELMIWLNMGVWPGGSRVATVELAGA-----TWEEVYADMDW 158  
DB 217 S--TSSGSSY--GSMVFSPTFMMSDMSAGONAMNOAAVATLTFYFAPKNIMVLAYQMSW 273  
QY 159 NYIAYRRITPT 169  
DB 274 PTLYRTSSDPT 284

RC STRAIN-K12;  
RX MEDLINE-97061202; PubMed-8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Ikeda T., Itoh T., Kajiura M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map."  
RL DNA Res. 3:137-155(1996).  
CC -----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL: AE000172; AAC73775.1; -  
CC DR EMBL: D90707; BAA35329.1; ALT\_INIT.  
CC DR EcoGene: EG13659; YbFM.  
CC KW Hypothetical protein: Complete proteome.  
CC SQ SEQUENCE 468 AA; 52780 MW; 954B5A778A1C2E4 CRC64;

Query Match 6.7%; Score 83; DB 1; Length 468;  
Best Local Similarity 23.6%; Pred. No. 4.5;  
Matches 46; Conservative 25; Mismatches 70; Indels 54; Gaps 10;

QY 44 RADHDNGNNVAAYPAIFGCGHMGACTNSGLPRVQELSDVRSWT----- 89  
DB 281 RSYNDYDGTAWLOALTFG-----YRADVDRLREGWVAKDGOOGYFLQR 327  
QY 90 LTPITGRMAAVIDMFSPVTSNGNGYSGAELMIML-----NNNGVWPGGSRVATV 142  
DB 328 MTP-TYASSNGRLDIM---DNRSDFNAGEKAVFFGAMYDLKNN---LPFAAGASY 379  
QY 143 ELGATWEVYADMDNYIAY--RRTPTTSVSELDKAEIDDAVARGYIRPMYLHAYE 200  
DB 380 VYA---WDAPKATWQSPDAYDYKNTRIEESAVSLDAVYIQGRAKGTW---FKLHETE 433  
QY 201 ----TGFELEGGAG 211  
DB 434 YDNHSDIPSWGGYG 448

RESULT 10  
YBFM\_ECOLI  
ID YBFM\_ECOLI STANDARD: PRT; 468 AA.  
AC P75733;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein ybfm.  
GN YBFM OR B0681.  
OS *Escherichia coli*.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC *Escherichia*.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
RX MEDLINE-97426617; PubMed-9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Berna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirtpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12."  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.

RESULT 11  
GUXB\_CELFI  
ID GUXB\_CELFI STANDARD: PRT; 1090 AA.  
AC P50899;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Exoglucanase B precursor (EC 3.2.1.91) (Exocellulohydrolase B)  
DE (1,4-beta-cellulohydrolase B) (CBP120).  
GN CBHB OR CENE.  
OS Cellulomonas fami.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.  
OX NCBI\_TaxID=1708;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.  
RC STRAIN-ATCC 484;  
RX MEDLINE-96003898; PubMed-7575482;  
RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;  
RT "Cellulohydrolase B, a second exo-cellulohydrolase from the  
RT cellulytic bacterium *Cellulomonas* fami.",  
RL Biochem. J. 311:67-74(1995).  
RN [2]  
RP SEQUENCE OF 54-75.  
RX MEDLINE-93209933; PubMed-8458833;





FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 819 819 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 840 840 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 921 921 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1200 1200 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1294 1294 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1311 1311 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1358 1358 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1449 AA; 159916 MW; PFC3CB8AC9FB655 CMC64;

Query Match 6.7%; Score 82.5; DB 1; Length 1449;  
Best Local Similarity 24.6%; Pred. No. 17;  
Matches 31; Conservative 13; Mismatches 43; Indels 39; Gaps 6;

QY 63 CHNGA-----CTNSGLPRRQGLSDVTSWTLPTTGRMAADI-WESPVTN 111  
DB 148 CMMGSECLNHRKPPICPSN-----EANGCMMLGLQMFADAVY 186  
QY 112 SGNGYSGGAEIMTWL--NNGGVMPGGSRYATVELAGATWEVYAD--WDMNYIAYRRRT 167  
DB 187 A---YILHGASYSRISFENMGSTVTLGDMRAVTLTETAGTLVDLMWFNVYDVSRYVANKN 243  
QY 168 PTSTVS 173  
DB 244 GTTVVS 249

RESULT 13  
THTR.CORGL STANDARD; PRT; 225 AA.  
AC P71121;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Thiosulfate sulfurtransferase (EC 2.8.1.1) (Fragment).  
GN THTR.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;  
OC Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032;  
RX MEDLINE=96337861; PubMed=8772169;  
RA Jaeger W., Peters-Wendisch P.G., Kallnowski J., Puhler A.;  
RT "A.Corynebacterium glutamicum gene encoding a two-domain protein  
similar to biotin carboxylases and biotin-carboxyl-carrier proteins.";  
RL Arch. Microbiol. 166:76-82(1996).  
CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
CC -1- SIMILARITY: BELONGS TO THE RHODANSE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: U35023; AAB04889.1; -  
CC DR HSSP; P52197; 1E0C.  
DR InterPro: IPR001307; Rhodanese.  
DR InterPro: IPR001763; Rhodanese\_domain.  
DR Pfam: PF00581; Rhodanese.1.

DR SMART; SM00450; RHOD; 1.  
DR PROSITE: PS00380; RHODANSE\_1; PARTIAL.  
DR PROSITE: PS00683; RHODANSE\_2; 1.  
KW Transferase.  
FT NON-TER 1 1  
FT ACT\_SITE 172 172 BY SIMILARITY.  
SQ SEQUENCE 225 AA; 25085 MW; B15A552DFCABA95B CRC64;

Query Match 6.6%; Score 81.5; DB 1; Length 225;  
Best Local Similarity 27.3%; Pred. No. 2.7;  
Matches 44; Conservative 20; Mismatches 56; Indels 41; Gaps 9;

QY 1 ELGRMDARDVAGRGYRVINNVGAEIACIEVGLG--TGNF-TITRADHDGNNVAYP 57  
DB 36 ELFGHSDVRLNNGR-----DAMMAEERDSYVPEPSANTYVERVDENGRAPFAEV- 89  
QY 58 AIYFGCHWAGCTSNGLPRRQGLSDVTSWTLPTTGRMAADYIWESPVTNSG---N 114  
DB 90 -----LGSITQSGCW-----TLVDVRRPSEFSGL-----DEHGNPTSGTLRG 128  
QY 115 GYSGAEIMTWLNMNGGVMPGGSRYATVELAGATWEVYAD 155  
DB 129 GHIPGA---INLDMSDAVLPNGNFRTRAEI-----DKLTAD 161

RESULT 14  
VGL2\_CVPPR STANDARD; PRT; 1447 AA.  
ID VGL2\_CVPPR  
AC 002167;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein).  
GN S.  
OS Porcine transmissible gastroenteritis coronavirus (strain Pur46-MAD).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=33739;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92410651; PubMed=1326823;  
RA Sanchez C.M., Gebauer F., Sune C., Mendez A., Dopazo J., Enjanes L.;  
RT "Genetic evolution and tropism of transmissible gastroenteritis  
coronaviruses.";  
RL Virology 190:92-105(1992).  
CC -1- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS  
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION  
CC AND IN SYNCTIUM FORMATION.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: M94101; AAA47109.1; -  
CC DR InterPro: IPR002351; Corona\_S1.  
DR InterPro: IPR002352; Corona\_S2.  
DR Pfam: PF01600; Corona\_S1; 1.  
DR Pfam: PF01601; Corona\_S2; 1.  
KW Glycoprotein; Envelope protein; Transmembrane; Signal.  
FT SIGNAL 1 16  
FT CHAIN 17 1447  
FT DOMAIN 17 1388  
FT TRANSMEM 1389 1408  
FT DOMAIN 1409 1447  
FT DOMAIN 1409 1430  
FT DOMAIN 405 465  
FT CARBOHYD 26 26  
N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	42	42	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	94	94	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	243	243	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	250	250	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	285	285	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	334	334	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	345	345	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	362	362	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	403	403	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	447	447	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	514	514	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	530	530	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	552	552	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	592	592	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	702	702	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	723	723	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	778	778	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	817	817	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	832	832	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	838	838	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	919	919	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	1072	1072	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	1198	1198	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	1292	1292	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	1309	1309	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	1322	1322	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	1334	1334	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	1339	1339	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	1356	1356	N-LINKED	(GLCNAC . .)	(POTENTIAL)
FT	CARBOHYD	1369	1369	N-LINKED	(GLCNAC . .)	(POTENTIAL)
EQ	SEQUENCE	1447 AA:	160115 MW:	062EBD052DE26637	CRC64:	

Query Match	6.6%	Score 81.5	DB 1	Length 1447
Best Local Similarity	24.6%	Pred. No. 21		
Matches 31, Conservative 14; Mismatches 42; Indels 39; Gaps 6				

```

Oy 63 CHMG-----CPSNGLPRVQELSDVRSWLTPTTGRMAAAD--NESVTN 111
    |::|      |||
Db 148 CMMSECLNKHKFPICPNS-----EAACGMMLYGLQKFA---D 183

Oy 112 SGNGSGGAELMIWL--NMNGVMPGGSRVATVELAGATWEVYAD--WMNTAIYRRTT 167
    |::|      |::|      |::|      |::|
Db 184 EYVALHMASTRISFENOMSGTTFGDMRATLTLEVAGLYDLMFNEPVDSYIRVNNKN 243

Oy 168 PTTSVS 173
    ||  ||
Db 244 GTTVVS 249
    ||  ||

```

RESULT	15
VGL2_CVPPU	
ID	VGL2_CVPPU
STANDARD;	
PRT;	1447 AA

DT 01-AUG-1988 (Rel. 01, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE E2 glycoprotein precursor (Spike glycoprotein) (Peplover protein).  
GN 5.  
OS Porcine transmissible gastroenteritis coronavirus (strain Purdue).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=1151;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=87253116; PubMed=3037011;  
RA Raaschaert D., Laude H.;  
RT "The predicted primary structure of the peplomer protein E2 of the  
RT porcine coronavirus transmissible gastroenteritis virus.";  
RL J. Gen. Virol. 68:1883-1890(1987).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=88129049; PubMed=2829461;  
RX

[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2002, 10:57:39 ; Search time 45.43 Seconds

(without alignments)  
845,364 Million cell updates/sec

Title: US-10-003-759-2\_COPY\_40\_261

Perfect score: 1234  
Sequence: 1 ELGCRMDARDVAGGRYRVIN.....FELMEGAGLSADFSVTYQ 222

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_RVIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1106.5	89.7	260	2	033897 Rhodothermu
2	396.5	32.1	371	2	09KIH1 streptomyc
3	352	28.5	382	2	059963 streptomyc
4	348.5	28.2	377	2	008468 streptomyc
5	345	28.0	381	2	09RXY3 streptomyc
6	343	27.8	381	2	054331 streptomyc
7	333	27.0	384	2	09X602 streptomyc
8	242.5	19.7	264	2	031030 streptomyc
9	204.5	16.6	244	2	09KXSA streptomyc
10	200.5	16.2	274	16	060033 thermotoga
11	196.5	15.9	274	2	P96492 thermotoga
12	194	15.7	151	16	053438 mycobacteri
13	178.5	14.5	238	3	094218 aspergillus
14	176.5	14.3	234	3	000095 trichoderma
15	172.5	14.0	239	3	013454 aspergillus
16	161.5	13.1	239	3	074705 aspergillus

17	160.5	13.0	319	1	09V2T0 pyrococcus
18	156.5	12.7	263	2	09AN41 bradyrhizob
19	153	12.4	257	2	008428 thermotoga
20	153	12.4	257	2	060032 thermotoga
21	153	12.4	257	2	P96491 thermotoga
22	153	12.4	258	16	09S5X8 thermotoga
23	142.5	11.5	332	17	09YXG7 sulfolobus
24	139	11.3	334	17	097X08 sulfolobus
25	136	11.0	239	3	09P8N6 caldicellul
26	106.5	8.6	361	2	052375 caldicellul
27	98	7.9	608	2	093GM7 sulfolobus
28	95.5	7.7	322	17	09YV57 penicillium
29	94.5	7.7	516	3	09URR7 penicillium
30	94.5	7.7	516	3	09URR6 penicillium
31	94	7.6	112	5	09V6J2 drosophila
32	93.5	7.6	673	16	007178 mycobacteri
33	92.5	7.5	840	10	093X57 fragaria an
34	91.5	7.4	690	5	09XWC5 caenorhabd
35	91	7.4	950	16	09A4B3 caulobacter
36	90.5	7.3	516	3	09URR5 penicillium
37	89.5	7.3	574	2	086876 streptomyc
38	89.5	7.3	1287	2	093105 burkholderi
39	89	7.2	242	5	096687 lumbricus b
40	88	7.1	928	10	09LTV0 arabidopsis
41	87	7.1	604	17	09HRM4 halobacteri
42	86.5	7.0	360	2	P77853 dicytyolomu
43	86.5	7.0	457	2	052779 clostridium
44	86.5	7.0	683	2	087119 clostridium
45	86	7.0	842	2	09XCV4 cellulomona

#### ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	260 AA.
033897	033897	033897		
AC	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	CELLULOSE (EC 3.2.1.4).			
GN	CELA.			
OS	Rhodothermus marinus (Rhodothermus obamensis).			
OC	Bacteria; CFB group; Rhodothermus group; Rhodothermus.			
OX	NCBI_TaxID=29549;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATN-TT1378;			
RX	MEDLINE=98242392; PubMed=9581291;			
RA	Hallidorsdottilir S., Thorolfssdottilir E.T., Spilliaert R., Johansson M.,			
RA	Thorbjarnardottilir S.H., Palsdottilir A., Hreggvidsson G.O.,			
RA	Kristjansson J.K., Holst O., Eggertsson G.;			
RT	Cloning, sequencing and overexpression of a Rhodothermus marinus gene			
RT	encoding a thermostable cellulase of glycosyl hydrolase family 12.7;			
RL	Appl. Microbiol. Biotechnol. 49:277-284 (1998).			
DR	EMBL: U72637; AAB55594.1;			
DR	InterPro: IPR002594; Glyco_hydro.12.			
DR	Pfam: PF01670; Glyco_hydro.12; 1.			
DR	Prodom: PD004316; Glyco_hydro.12; 1.			
KW	Hydrolase; Glycosidase.			
SQ	SEQUENCE 260 AA; 28770 MW; 94F197DB7D0D247A CRC64;			

Query Match 89.7%; Score 1106.5; DB 2; Length 260;

Best local Similarity 88.6%; Pred. No. 3.3e-87;

Matches 202; Conservative 4; Mismatches 9; Indels 13; Gaps 2;

QY 1 ELGCRMDARDVAGGRYRVINNVGAETACIEVGLTGTTRADHDGNNVAAYPAIY 60  
Db 40 ELGCRMDARDVAGGRYRVINNVGAETACIEVGLTGTTRADHDGNNVAAYPAIY 99  
QY 61 FGCHMGA-----CTNSGLPRVQELSDVTRTSWTLPTITGMMNAAYIWPSPVNSGN 114

```
|||||
: : : : :
Db 100 FCGHMAPARAIIDCAACAGACVRAHELD-----VTPTTGRMNAAYDWFSPVNSGN 152
115 GYSGGALMIWLMWNGVWGSGSRVATVELAGATWEVYADWMNYIAVRRTPTTSVSE 174
153 GTSGGALMIWLMWNGVWGSGSRVATVELAGATWEVYADWMNYIAVRRTPTTSVSE 212
QY 175 LDKAFIDDAVARGYIRPEWYLHAVETGELMEGAGLSRADSFSVTVQ 222
|||||
Db 213 LDKAFIDDAVARGYIRPEWYLHAVETGELMEGAGLSRADSFSVTVQ 260

RESULT 2
ID 09KIH1 PRELIMINARY: PRT; 371 AA.
AC 09KIH1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULOSE 12A.
GN Streptomyces sp. 11A68.
OS Streptomyces sp. 11A68.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1334352;
RN [1]
RC STRAIN=11A68;
RA van Solingen P., Meijer D., van der Kleij W.A.H., Barnett C.C.,
RA Bolle R., Power S.D., Jones B.E.;
RT "Cloning and expression of an endocellulase gene from a novel
RT Streptomyces isolated from an East African soda lake.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF233376; AAF91283.1; -
DR HSSP; P07986; 1EXG.
DR InterPro; IPR002594; Glyco_hydro_12.
DR InterPro; IPR001230; Prenyltn.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
SQ SEQUENCE 371 AA; 38481 MW; 0E1BC428BA148914 CRC64;

Query Match 32.1%; Score 396.5; DB 2; Length 371;
Best Local Similarity 39.7%; Pred. No. 3.9e-26;
Matches 89; Conservative 27; Mismatches 99; Indels 9; Gaps 6;

QY 1 ELGGRDADVAGRRYRINNWGCAETACIEVGLTGN-FTTRADHNGNVA--AYP 57
: : : : :
Db 34 QICDRYGTITIQ-DRYVONNNRNGTSATOCIN--TGNGFETQADGSVPTGAPKRSYP 89
58 AIFYGCHMGACTNSGLPRVQELSDVRTSWTLPTTTGRMNAAYDIWFSPVNSGNGYS 117
: : : : :
Db 90 SYVDGCHYGCAPRTTLPRHISISIGASPSVSTRYNGNGVYNAAYDIWIDP-TPRINGVN 148
118 GGAELMIWLMWNGVWGSGSRVATVELAGATWEVYADWMNYIAVRRTPTTSVSELDL 177
: : : : :
Db 149 -RTEIMWTFNRVGPYQIPISPVGTAVHVGSRSEWVTGNSGNSDVISFLAPSAISSMSFDV 207
178 KAFIDDAVARGYIRPEWYLHAVETGELMEGAGLSRADSFSVTV 221
: : : : :
Db 208 KDFVDVAVSHGLATPDWYLTISIQAGFEPWGGTGLAVNSSFSAV 251

RESULT 3
ID 059963 PRELIMINARY: PRT; 382 AA.
AC 059963;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELLULOSE (EC 3.2.1.4).
GN EGLS.
```

```
OS Streptomyces rochei (Streptomyces parvulus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1928;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=A2;
RX MEDLINE=95011642; PubMed=7523249;
RA Perito B., Hanhart E., Irdani T., Iqbal M., McCarthy A.J.,
RA Mastromel G.;
RT "Characterization and sequence analysis of a Streptomyces rochei A2.";
RL Gene 148:119-124(1994).
DR EMBL; X73953; CAA52139.1; -.
DR HSSP; P07986; 1EXG.
DR InterPro; IPR001919; CBD_2.
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF00553; CBD_2; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
DR PROSITE; PS00561; CBD_BACTERIAL; UNKNOWN_1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 382 AA; 39398 MW; 21C014342EFC6565 CRC64;

Query Match 28.5%; Score 352; DB 2; Length 382;
Best Local Similarity 36.2%; Pred. No. 2.6e-22;
Matches 77; Conservative 32; Mismatches 98; Indels 6; Gaps 4;

QY 11 VAGGRYRINNWGCAETACIEVGLTGNFTTRADHNGNVA--AYPAFYGCHMGAC 68
: : : : :
Db 49 VIGGRYVONNNRNGTSATOCV-TATDSG-FRVTQADGSPVTPNAPKRSYVFNGCHITNC 106
69 TNSGGLPRVQELSDVRTSWTLPTTTGRMNAAYDIWFSPVNSGNGYSGGALMIWLMW 128
: : : : :
Db 107 SPFTALPARISGISAPSSSYGFVNAYNASTYDILMDPTPTDG--VNRTEIMWFRN 164
129 NGCVWPGSGSRVATVELAGATWEVYADWMNYIAVRRTPTTSVSELDKAFIDDAVARG 188
: : : : :
Db 165 VGIQIPITGSGVGTASVAGRTWEVWSGGNGTNDVLSFVARSAMSMSFEDVDFRATVARG 224
189 YIRPEWYLHAVETGELMEGAGLSRADSFSVTV 221
: : : : :
Db 225 LAGNDWYLTISIQAGFEPWONGAGLAVNSSFSAV 257

RESULT 4
ID 008468 PRELIMINARY: PRT; 377 AA.
AC 008468;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CEL2 (EC 3.2.1.4).
GN CEL2.
OS Streptomyces halstedii.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1944;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=JMB;
RX MEDLINE=97307849; PubMed=9182697;
RA Garda-Salas A.L., Fernandez-Abalos J.M., Sanchez P., Ruiz-Arribas A.,
RA Santamaria-Sanchez R.I.;
RT "Two genes encoding an endoglucanase and a cellulose-binding protein
RT are clustered and co-regulated by a TTA codon in Streptomyces
RT halstedii JMB.";
RL Biochem. J. 324:403-411(1997).
DR EMBL; U51222; AAC45429.1; -.
DR HSSP; P07986; 1EXG.
DR InterPro; IPR001919; CBD_2.
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF00553; CBD_2; 1.
```

DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
DR Prodom: PD004316; Glyco\_hydro\_12; 1.  
KW Hydrolase; Glycosidase.  
SQ SEQUENCE 377 AA; 38829 MW; 9892191B7A1EC280 CRC64;

Query Match 28.2%; Score 348.5; DB 2; Length 377;  
Best Local Similarity 35.9%; Pred. No. 5.2e-22;  
Matches 80; Conservative 43; Mismatches 91; Indels 9; Gaps 8;

QY 2 LGRMDARADVAGRYRINNVWGAEETAOCEIEVGLTGNFTITRADHNGNVA--AYPAI 59  
DB 41 VCGQYSTTIQ-GRVYQNNRMGASAPQCV-TATDSG-FRVYQADGAPTPNGAPKSPV 97  
QY 60 YFGCHGACTSNGLPRVQELSDVTSWTLPTTGRMNAAYDIFSPVTSNGNGYSGG 119  
DB 98 FNGCHTNCSPGNTLPAQVSGIASAPSSISYGFVGSAYVNASYDIDLDP-TPKKNGN-R 155  
QY 120 AELMILNMGVMPGSGRVAIVELAGATWEVYADMDN-YIAYRRTPTTSVSELDK 178  
DB 156 TEIMIMLNKVGPIQPIGSGAGTASVGRTOVVRGNSGNDVISFVAPSAVASMS-FDVM 214  
QY 179 AFIDDAVARGIPEWYLHVENGFELMEGAGLRSDAFSVTV 221  
DB 215 DFVRNTIARGMAQNNMYLTISVQGFEPQNGAGLAVNSFSSTV 257

## RESULT 5

Q9RJY3 PRELIMINARY; PRT; 381 AA.

AC 09RJY3; 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PUTATIVE SECRETED CELLULOSE B.  
GN CELB.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Oliver K., Harris D.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN 12  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
RN 13  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Redenbach M., Krieser H.M., Denapate D., Eichner A., Cullum J.,  
RA Kinashchi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL EMBL: AL133210; CAB61599.1; -.  
DR HSSP: P07986; IEXG.  
DR InterPro: IPR001919; CBD\_2.  
DR InterPro: IPR002594; Glyco\_hydro\_12.  
DR Pfam: PF00553; CBD\_2; 1.  
DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
DR Prodom: PD004316; Glyco\_hydro\_12; 1.  
DR PROSITE: PS00561; CBD\_BACTERIAL; UNKNOWN\_1.  
SQ SEQUENCE 381 AA; 39199 MW; 97CD8F58679E4EC CRC64;

Query Match 28.0%; Score 345; DB 2; Length 381;  
Best Local Similarity 36.5%; Pred. No. 1e-21;  
Matches 77; Conservative 31; Mismatches 97; Indels 6; Gaps 4;

QY 14 GRRVYNNVWGAEETAOCEIEVGLTGNFTITRADHNGNVA--AYPAIFGCHGACTSN 71  
DB 55 GRRVYNNRMGSGTATQCV-TATDTG-FRVYQADGASAPTPNGAPKSPSVFNGCHYTCNSPG 112  
QY 72 SGLPVRVQELSDVTSWTLPTTGRMNAAYDIFSPVTSNGNGYSGGAEIMTLMWNG 131  
DB 113 TALPVRIDTVSAAPSSISYGFVDAVYNASYDIDLDPARTDG--VNOETIMTFNRVGP 170  
QY 132 VMEGSRVATVELAGATWEVYADMDNRYIAYRRTPTTSVSELDKAFIDDAVARGYIR 191  
DB 171 IOPIGSPVGTASVGRTEWPMGSGNSNDVLSFVAPSAISGNSFDVMDFRATVARGLA 230  
QY 192 PEWYLHVENGFELMEGAGLRSDAFSVTV 222  
DB 231 NDVYLTISVQGFEPQNGAGLAVNSFSSTV 261

## RESULT 6

Q54331 PRELIMINARY; PRT; 381 AA.

AC 054331; 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE CELLULOSE B PRECURSOR.  
GN CELB.  
OS Streptomyces lividans.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1916;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN-66;  
RA Wittmann S., Sharek F., Kluepfel D., Morosoli R.;  
RT "Purification and characterization of the CelB endoglucanase from  
RT Streptomyces lividans 66 and DNA sequence of the encoding gene.";  
RL Appl. Environ. Microbiol. 60:1701-1703(1994).  
RN 12  
RP SEQUENCE FROM N.A.  
RC STRAIN-66;  
RA Sharek F.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U04629; AAB71950.1; -.  
DR HSSP: P07986; IEXG.  
DR InterPro: IPR001919; CBD\_2.  
DR InterPro: IPR002594; Glyco\_hydro\_12.  
DR Pfam: PF00553; CBD\_2; 1.  
DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
DR Prodom: PD004316; Glyco\_hydro\_12; 1.  
DR PROSITE: PS00561; CBD\_BACTERIAL; UNKNOWN\_1.  
KW Signal.  
FT SIGNAL 1 40  
FT CHAIN 41 381 CELLULOSE B.  
SQ SEQUENCE 381 AA; 39239 MW; A7E99BF590FA24EC CRC64;

Query Match 27.8%; Score 343; DB 2; Length 381;  
Best Local Similarity 36.5%; Pred. No. 1.5e-21;  
Matches 77; Conservative 31; Mismatches 97; Indels 6; Gaps 4;

QY 14 GRRVYNNVWGAEETAOCEIEVGLTGNFTITRADHNGNVA--AYPAIFGCHGACTSN 71  
DB 55 GRRVYNNRMGSGTAPQCV-TATDTG-FRVYQADGASAPTPNGAPKSPSVFNGCHYTCNSPG 112  
QY 72 SGLPVRVQELSDVTSWTLPTTGRMNAAYDIFSPVTSNGNGYSGGAEIMTLMWNG 131  
DB 113 TDLPRVLDTVSAAPSSISYGFVDAVYNASYDIDLDPARTDG--VNOETIMTFNRVGP 170  
QY 132 VMEGSRVATVELAGATWEVYADMDNRYIAYRRTPTTSVSELDKAFIDDAVARGYIR 191  
DB 171 IOPIGSPVGTASVGRTEWPMGSGNSNDVLSFVAPSAISGNSFDVMDFRATVARGLA 230

OY 192 PENYLAHETGFEWEGAGLRSDPSVTQ 222  
 DB 231 NDWYTSVQAGFEPQMGAGLAVNSFSSTV 261

RESULT 7  
 OYX602 PRELIMINARY: PRT: 384 AA.  
 ID O9X602;  
 AC O9X602;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE CELULYASE.  
 GN CELS.  
 OS Streptomyces viridosporus.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 NCBI\_TaxID=67581;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-T7A;  
 RA Ramachandran S., Crawford D.L.;  
 RT "Characterization and sequence analysis of two genes involved in  
 cellulase degradation in Streptomyces viridosporus T7A, and its  
 expression in *Escherichia coli*.";  
 RT Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AF130408; AAD25090.1; -  
 DR HSSP: P07986; 1EXG.  
 DR InterPro: IPR001919; CBD\_2.  
 DR InterPro: IPR002594; Glyco\_hydro\_12.  
 DR Pfam: PF00553; CBD\_2; 1.  
 DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
 DR ProDom: PD004316; Glyco\_hydro\_12; 1.  
 SQ SEQUENCE 384 AA; 40918 MW; D3968BBE8DEDE65 CRC64;

Query Match 27.0%; Score 333; DB 2; Length 384;  
 Best Local Similarity 35.0%; Pred. No. 1.1e-20;  
 Matches 77; Conservative 34; Mismatches 93; Indels 16; Gaps 6;

OY 14 GRVRVNNWGAETACIEVGLTGNFTTIRADHNCNNVA--AYAIYGCWMACTSN 71  
 DB 52 GRVYVNNWNRGSSPCCV-TATDTG-FRLQADGVPPTNAPKSYBSVFNCHYTNGSPG 109  
 OY 72 SGLPRRVOELSDVRT-----SWTLFTPTTGR-----WNAAYDWFSPVTSNGYSGAG 121  
 DB 110 TKLPARISGSSAPARITXGSSAPSSISYGVGAVYMASYDMLDPTPTDQ--VNRT 167  
 OY 122 LMIWLNMGVMPGSRVATVELAGATWEVYVADMDMNYIAYRRTPPTTSVSELDLKAFT 181  
 DB 168 IMIMFKKVGPIQPIGSGVGTATVCGRTQWVSGNGSNDLSEFVAPSAIESMGFDVDMFV 227  
 OY 182 DDAVARGYIRPEWYLAHETGFEWEGAGLRSDPSVTQ 221  
 DB 228 RETVARGMAQNDWYLTYSVQAGFEPQMGAGLAVNSFSSTV 267

RESULT 8  
 OYX602 PRELIMINARY: PRT: 264 AA.  
 ID O31030;  
 AC O31030;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)  
 DE BETA(1,4)-GLUCAN GLUCANOHYDROLASE PRECURSOR.  
 GN CELB.  
 OS Pectobacterium carotovorum subsp. carotovorum.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Pectobacterium.  
 NCBI\_TaxID=555;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-LY34;  
 RA MEDLINE=98096373; PubMed=9434760;  
 RX Park Y.W., Lim S.T., Cho S.J., Yun H.D.;  
 RT "Characterization of *Erwinia carotovora* subsp. *carotovora* LY34 endo-  
 1,4-beta-glucanase genes and rapid identification of their gene  
 products.";  
 RT Biochem. Biophys. Res. Commun. 241:636-641(1997).  
 RL EMBL: AF025769; AAC02965.1; -  
 DR InterPro: IPR002594; Glyco\_hydro\_12.  
 DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
 DR ProDom: PD004316; Glyco\_hydro\_12; 1.  
 KM Signal; Hydrolase.  
 FT SIGNAL 1 36  
 FT CHAIN 37 264  
 FT SEQUENCE 264 AA; 29634 MW; 70EA3666B8443CEB CRC64;

Query Match 19.7%; Score 242.5; DB 2; Length 264;  
 Best Local Similarity 29.8%; Pred. No. 4.1e-13;  
 Matches 68; Conservative 39; Mismatches 90; Indels 31; Gaps 9;

OY 15 RYRVNNWGAETACIEVGLTGNFTTIRAD-----HDGNNVAAYPAIYGCWMA 67  
 DB 48 KYVFNWNGKDEYK-----GMQOTVEYNSPTSMGMMHWPSSSYKATPSLYSGHMTA 103  
 OY 68 C-TSNGSLPRRVOELSDVRTSWTLFTPTTGRWNAAYDIWFSPVTSNGYSGAELMIWL 126  
 DB 104 VYTENSGLPFKLSNKSITSNVYSIKSFATLYAAVDVWFHTDKASMDSTPDELMIWL 163  
 OY 127 NNWGVMPGSRVATVELAGATWEVYVADY-----DMWYIAYRRTPPTTSVSELDLKA 179  
 DB 164 N-NTNAGPADYLETVELDGSSMNV--FKGINAGNGGMMVFSFVPTSNAS-LNIRH 220  
 OY 180 FIDDAVARGYIRPEW-----YLAHETGFEWEGAGLRSDPSVTQ 222  
 DB 221 FTNDVQ-----RKKWMSNAKISSVELGTETIFGDDQIDITKNSVDK 264

RESULT 9  
 OYX54 PRELIMINARY: PRT: 244 AA.  
 ID O9KX54;  
 AC O9KX54;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE PUTATIVE SECRETED SUGAR HYDROLASE.  
 GN SC5H4.15.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Murphy L., Harris D.;  
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denaplatte D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL355913; CAB91127.1; -  
 DR InterPro: IPR002594; Glyco\_hydro\_12.  
 DR Pfam: PF01670; Glyco\_hydro\_12; 1.



DR	PfDomId:	PD004316;	Glyco_hydro_12;1.
KW	Hydrolase.		
SO	SEQUENCE	244 AA;	26123 MW; 2CA63242798BEND44 CRC64;
	Query Match	16.6%;	Score 204.5; DB 2; Length 244;
	Best Local Similarity	26.7%;	Pred. No. 6,7e-10;
	Matches 60;	Conservative 52;	Mismatches 78; Indels 35; Gaps 11;
OY	3	CGRDARDVAGRRRVYNWGAET-AOCI--ENGLETGNTTTRADHDNGNNVAAYPAI	59
	I : I :	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :	
Db	46	CDQWGNTSLNC--YLTYLNNISGWCSAGSCVANNSTGDVGW---ADHPNRTGIKSP--	97
OY	60	YFGCHWGCCTSNSSGLPRVVOELSDVRTSMITLPITTGRRNNAYIDWFSPYTNMGNSGSG	119
	I : I :	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :	
Db	98	-----NAKKVIINKRITSLSLTSSYNVTVPSSGAYNTSYDIW----DTDYD---	140
OY	120	AELMIWLNMGNCGVMPGSGSRVATVELAGATWEWYADMOMNYI-AVRRTPTTS--VSELD	176
	I : I : I :	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :	
Db	141	EIMLVNVHHAAGVLGTGGGSVGLGGHTTWYDKKNNGANEFSLRLIRSDSNSGYINILP	199
OY	177	LKAFLDDAVANGYIRPEWYLHAVETGFELMEGGAGLRSADESVTV	221
	I : I :	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :	
Db	200	ILKWIKD--TKGMWGNE-TIGDVQGEYEITSSAGGLDPRTNLTV	241
	I : I :	I : I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :	
RESULT	10		
ID	060033	PRELIMINARY;	PRF; 274 AA.
AC	060033;		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DF	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	
DE	ENDO-1,4-BETA-GLUCANASE PRECURSOR (EC 3.2.1.4) (CELLULASE)		
DN	(ENDOGLUCANASE) (CARBOXYMETHYL CELLULASE).		
OS	Thermotoga maritima.		
OC	Bacteria; Thermotogales; Thermotoga.		
RX	NCHI_TaxID=2336;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MSB8 (DSM3109);		
RX	MEDLINE=96425879; PubMed=8828221;		
RA	Liebl W., Rulle P., Bronnenmeier K., Riedel K., Lottspeich F.,		
RT	Greif I.;		
RT	"Analysis of a Thermotoga maritima DNA fragment encoding two similar		
RT	thermostable cellulases, Cella and CelB, and characterization of the		
RT	recombinant enzymes.";		
RL	Microbiology 142:2532-2542(1996).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MSB8 / DSM 3109;		
RX	MEDLINE=99287316; PubMed=10360571;		
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,		
RA	Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,		
RA	McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,		
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,		
RA	Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,		
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;		
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from		
RT	genome sequence of Thermotoga maritima."		
RL	Nature 399:323-329(1999).		
DR	EMBL: Z69341; CAAS3274.1; -		
DR	EMBL: AE001800; AAD36592.1; -		
DR	TIGR: TM1525; -		
DR	InterPro: IPR002594; Glyco_hydro_12.		
DR	Pfam: PF01670; Glyco_hydro_12; 1.		
DR	PfDomId: PD004316; Glyco_hydro_12; 1.		
KW	Signal; Hydrolase; Glycosidase; Complete proteome.		
FT	SIGNAL	1	POTENTIAL.
FT	CHAIN	18	274
FT	SEQUENCE	274 AA;	31734 MW; 7200A5ACAD7BA358 CRC64;

```

QY      6 WDADVAGGRYRVINNWGAETACIEVELETGNETTIRADHDNGNNVAAYPAIFYGCH- 64
Query Match          16.2%; Score 200.5; DB 16; Length 274;
Best Local Similarity 27.9%; Pred.No.1,7e-09;
Matches 70; Conservative 39; Mismatches 85; Indels 57; Gaps 15;
Db       Db
Dy      65 MGACTSNG---LRRRQELSD--VRSWTL-----PTITGRKMAAVIDMFSPYTNSGN 114
        | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
        44 WNVSVSEEG-----EPLKFDGEKEVFEDLYLVNLQNPD----SWHGPEIYYGKP 92
        + + + + + + + + + + + + + + + + + + + + + + + + + + + +
OY      115 GYSGCAELMIWLNMNGVGPGGSRY---ATVELAG-----ATWEVYADMOMNTIARYRT 166
        | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       146 VSSGDALIMWF-YNNVLMPGGOKYDETTTVIELIGVAOEIKMDIVFAPMGNDILAFRLT 204
        + + + + + + + + + + + + + + + + + + + + + + + + + + + +
QY      167 TPPTSVS-ELDLKAFIDA---VARGYIRPBYLHAVETGFELMEGSA----- 210
        || : : : : : : : : : : : : : : : : : : : : : : : : : :
        205 TPMKEKKIKINVKDFYOKAEVVAKKHSTRID--NFEELYCVMWELGTFGDPNPTAAKF 261
        | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       Dy      211 GLRSADFSTVY 221
        | : : : : : : : : : : : : : : : : : : : : : : : : : :
        262 GWTFRDFSVEY 272

RESULT 11
P96492 ID P96492 PRELIMINARY; PRT; 274 AA.
AC P96492:
DT 01-MAY-1997 (TREMBLrel_03, Created)
DT 01-JAN-1998 (TREMBLrel_05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel_17, last annotation update)
DE ENDO-1,4-BETA-GLUCONASE B (EC 3.2.1.4) (CELLULASE) (ENDOGLUCCANASE)
DE (CARBOXYMETHYL CELLULOSE).
CEB.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_Taxid=2337;
[1]
RP SEQUENCE FROM N.A.
RA Bok J.D., Eveleigh D.E.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE OF 1-137 FROM N.A.
RC STRAIN=Z2706-MC24;
RA Zverlov V.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 1-137 FROM N.A.
RC STRAIN=Z2706-MC24;
RX MEDLINE=93356813; PubMed=8352795;
RA Dakhova O., Kuregina N., Zverlov V., Svetlichnyi V.,
   Velikodvorskaya G.;
RT Cloning and expression in Escherichia coli of Thermotoga neapolitana
   genes coding for enzymes of carbohydrate substrate degradation." ;
RL Biochim. Biophys. Res. Commun. 194:1359-1364(1993).
CC -I- CATALYTIC ACTIVITY: ENDODHROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
   LINKAGES IN CELLULOSE.
DR EMBL; U93354; AAC95060.1; -;
DR EMBL; Z86103; CAB06782.1; -;
DR InterPro; IPR002594; Glyco_hydro_12.
DR Pfam; PF01670; Glyco_hydro_12; 1.
DR ProDom; PD004316; Glyco_hydro_12; 1.
KW Hydrolase; Glycosidase.
SO SEQUENCE 274 AA; 31752 MW; 6B21CAF8CA7C9AA2 CRC64;
```

```
Db 44 WNAKSYEG-----ETWLKDEGEKVOEYADIVNYVLONPD-----SVNHGYPPEIYYGKRP 92
QY 65 WCACTSNSSG---LPRRVOELSD---VRTSWTL-----TPITGRMNAAYDIMEFSPVTSNGN 114
Db 93 WAA--HNSGTEILPVRKVKDLPDFYVTLIDYSIWENDLPT-----NLAMEWITRRKPPOTS 145
QY 115 GYSGAGELMIWLMWNGVMDGSGSRV---ATVELAGA---TWEEVYADMDMYIAVART 166
Db 146 VSSGDEIVWYMF--YNNILMPGSGKQDEFTTIEINSPEVETKNDVYFAPMGWIDYLAFLRLT 204
QY 167 TPTTSSV--ELDLAFIDDA---VARGTIREPWLHAVETGFELMEGA-----210
Db 205 TPKKDGKRVKRVNDFEKAELVIRKSHTRAVE---NPEDEMYCWEIGETGERGPNTTAKF 261
QY 211 GLRSADFSTV 221
Db 262 GMTFKDFSEVEI 272

RESULT 12
053438 PRELIMINARY: PRT: 151 AA.
AC 053438;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE HYPOTHEICAL 16.4 KDA PROTEIN.
RV1090 OR MTV017.43.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H378V;
MEDLINE=98293987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne K., Quail M.A., Squares S., Squares R.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998)
DR EMBL: AL021897; CAI17206.1; -
DR Tuberculise; RV1090; -
DR InterPro: IPR002594; Glyco_hydro.12.
DR Pfam: PF01670; Glyco_hydro.12; 1.
DR Prodom: PD004316; Glyco_hydro.12; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 151 AA; 16404 MW; 8BCC10CF9C8E54 CRC64;

Query Match 15.7%; Score 194; DB 16; Length 151;
Best Local Similarity 33.3%; Pred. No. 3e-09;
Matches 51; Conservative 16; Mismatches 78; Indels 8; Gaps 4;
```

```
RESULT 13
094218 PRELIMINARY: PRT: 238 AA.
ID 094218;
AC 094218;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE XYLOGLUCAN-SPECIFIC ENDO-BETA-1,4-GLUCANASE PRECURSOR.
OS Aspergillus aculeatus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5053;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KSM 510;
MEDLINE=99102417; PubMed=9884411;
RA Pauly M., Andersen L.N., Kauppinen S., Kofoed L.V., York W.S.,
RA Albersheim P., Darvill A.;
RA "A xyloglucan-specific endo-beta-1,4-glucanase from Aspergillus
RT aculeatus: expression cloning in yeast, purification and
RT characterization of the recombinant enzyme."
RL Glycobiology 9:93-100(1999).
DR EMBL: AF043595; AAD02275.1; -
DR InterPro: IPR002594; Glyco_hydro.12.
DR Pfam: PF01670; Glyco_hydro.12; 1.
DR Prodom: PD004316; Glyco_hydro.12; 1.
KW Signal.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 238 XYLOGLUCAN-SPECIFIC ENDO-BETA-1,4-
FT GLUCANASE.
SQ SEQUENCE 238 AA; 25158 MW; FCCA6746D9AEC1B1 CRC64;
```

```
Query Match 14.5%; Score 178.5; DB 3; Length 238;
Best Local Similarity 25.4%; Pred. No. 1.1e-07;
Matches 63; Conservative 39; Mismatches 89; Indels 57; Gaps 12;

QY 1 ELGRMDAROVAGRRVINVWGAET--AOCIEVGLTGNFTTRADHGNNAVAP 57
Db 22 DFCGQMDT--ATAGDFTLYNDLMGESAGTGSQC-----TGVDSYSGDTIA--- 64
QY 58 AITFGCHWCACTNSGLPRRVOELSDVRSWTLPTI-----TTGRN-----N 99
Db 65 -----WHTSMWSGSGSSSVK--SYVNALITPTPOLNCISSIPTTWKSYSGSSIVAD 115
QY 100 AAYDIMEFSPVTSNGYSGSAGELMIWLMNNGVMP---GSSRVATVELAGATVEVYA-D 155
Db 116 VAYDTFLAETASGSSKY---ETVMWLALGGAGPISSTGSTITATPTTGVNKKLISGRN 171
QY 156 WDNVYIARRRTPTTSVSELDLKAFLDDAVANGYIRPEWYLHAVETGFELMES-GAGLRS 214
Db 172 GDTTVVSFVADSTTESFSG-DLNDFTYLVNDCGVDELITLLEAGTEPFGSNKKLTV 230
QY 215 ADFSVTV 222
Db 231 SEYSISIE 238

RESULT 14
000095 PRELIMINARY: PRT: 234 AA.
ID 000095;
AC 000095;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE ENDO-BETA-1,4-GLUCANASE (EC 3.2.1.4).
GN EGL.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
```

RP SEQUENCE FROM N.A.  
 RC STRAIN=Q99414;  
 RA Okada H., Tada K., Sekiya T., Yokoyama K., Takahashi A., Tonda H.,  
 RA Kumagai H., Morikawa Y.;  
 RT "Molecular characterization and heterologous expression of the gene  
 RT encoding a low-molecular-mass endoglucanase from *Trichoderma reesei*  
 RT QM9414."  
 RL Appl. Environ. Microbiol. 64:55-563(1998).  
 DR EMBL: AB003694; BAA20140.1;  
 DR InterPro: IPR002594; Glyco\_hydro\_12.  
 DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
 DR ProDom: PD004316; Glyco\_hydro\_12; 1.  
 KM Hydrolyase; Glycosidase.  
 SQ SEQUENCE 234 AA; 25159 MW; DF476EDE384ADD1 CRC64;

Query Match 14.3%; Score 176.5; DB 3; Length 234;  
 Best Local Similarity 28.6%; Pred. No. 1.6e-07;  
 Matches 68; Conservative 34; Mismatches 93; Indels 43; Gaps 14;

QY 3 CGRMDADVAGGRYRVINNWGAETAQ--CI-EVGLGTGFTITRAD--HDGNNVAA 55  
 DB 20 CDQW--ATFTNGYTVSNLMGASAGSGFCVAVSLSG--ASMHADQMWSGGQNNVKS 75  
 QY 56 YPAIYECGHGACTSNGLP--RRVQELSDVRT--SWTLPIITGRNNAAYDIWFSPTN 111  
 DB 76 YQ-----NSQIAIPQKRTVNSISSMPTTASWSYSG-SNIRANVAYDL-FTANP 122  
 QY 112 SCNGYSGAELMTLWLNNGGVMFGSRVATVELAGATWEVWY----ADMDNNTAYRRTT 167  
 DB 123 NHTYSGDYELMTLWLGKYGIDIGPIGSSQGTVNNGGQSWTLXGNGAMQYISFYAQTNTT 182  
 QY 168 PTTVSSELDIAKFI---DDAVARGYIRPEWYLAHVETGFELEWEGAGLRSADESVTV 221  
 DB 183 NYSG---DVKNFEFNYLRDN---KGYNAAGQYVLSYQFTEPFTGSGTLNVAWSWTASI 233

RESULT 15  
 ID 013454 PRELIMINARY; PRT; 239 AA.  
 AC 013454;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ENDO-1,4-BETA-GLUCANASE (EC 3.2.1.4).  
 GN CELA.  
 OS Aspergillus oryzae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillius.  
 OX NCBI\_TaxID=5062;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KBN616;  
 RX MEDLINE=97161783; PubMed=900887;  
 RA Kitamoto N., Go M., Shibayama T., Kimura T., Kito Y., Ohmiya K.,  
 RA Tsukagoshi N.;  
 RT "Molecular cloning, purification and characterization of two endo-1,4-  
 RT beta-glucanases from *Aspergillus oryzae* KBN616."  
 RL Appl. Microbiol. Biotechnol. 46:538-544(1996).  
 DR EMBL: D83731; BAA22588.1;  
 DR InterPro: IPR002594; Glyco\_hydro\_12.  
 DR Pfam: PF01670; Glyco\_hydro\_12; 1.  
 DR ProDom: PD004316; Glyco\_hydro\_12; 1.  
 KM Hydrolyase; Glycosidase.  
 SQ SEQUENCE 239 AA; 26096 MW; C0F850E5DFEB455D CRC64;

Query Match 14.0%; Score 172.5; DB 3; Length 239;  
 Best Local Similarity 27.3%; Pred. No. 3.6e-07;

Matches 63; Conservative 37; Mismatches 86; Indels 45; Gaps 11;

QY 1 ELGGRMDADVAGGRYRVINNWGAET--AOCIEVGLGTGFTITRADHDGNNVAAYP 57  
 ||| ::| : | | ||::| ::| ::| |

DB 18 ELCAQIDS--ASSPPYSVNNNLNGDSTGSCQCVY-----DNLSSSGA-- 59  
 QY 58 AIYFGCHW-----GACTSNGLPRRVQEL-SDVRTSWTLPIITGRNNAAYDI 104  
 DB 60 AMHTTWNNGEGSVKSYSNSAVTFPKKLSDVQSIPTDVMQODNTNVA---DVAIDL 116  
 QY 105 WFSPTVNSGNGYSGAELMTLWLNNGGVMFGSRVATVELAGATWEVWY----ADMDN 159  
 DB 117 -FTAADQNNHTYSGDYELMTLWLRGYTIQIDTATVEGHTEWELWGTIIQAGAEQK 175  
 QY 160 YIAYRRTPTTSVSELDIAKFIIDAVAR-GYIREWYLAHVETGFELEWEGG 209  
 DB 176 TYSFVSATPINTFEG-DIKKFDYITKHSFPASAOYLINMQFTEPFTG 225

Search completed: August 15, 2002, 10:57:39  
 Job time: 393 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 12:44:21 ; Search time 6219.51 Seconds  
(without alignments)  
2644.623 Million cell updates/sec

Title: US-10-003-759-3

Perfect score: 786  
Sequence: 1 atgaacgcatcgcgtgcggt.....ttccgtaacggtgcagtag 786

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_hlg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hlg\_hum: \*  
31: em\_hlg\_inv: \*  
32: em\_hlg\_other: \*  
33: em\_hlgo\_inv: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	786	100.0	786	6	AX339682	AX339682 Sequence
2	786	100.0	1877	6	AX339680	AX339680 Sequence
3	746.8	95.0	1874	1	RM072637	U72637 Rhodothermu
4	131.8	16.8	1116	6	AF233376	AF233376 Streptomy
5	131.8	16.8	1116	6	AR129926	AR129926 Sequence
6	131.8	16.8	1116	6	AR168360	AR168360 Sequence
7	105.8	13.5	1716	6	AR168362	AR168362 Sequence
8	97.8	12.4	3527	1	SH051222	U51222 Streptomyce
9	96.2	12.2	1470	1	SREGLS	X73953 S.rochei eg
10	89	11.3	1311	1	AF130408	AF130408 Streptomy
11	87.6	11.1	1565	1	SL004629	U04629 Streptomyce
12	87.6	11.1	41782	1	SCG11A	AL133210 Streptomy
13	69	8.8	11202	1	AE005123	AE005123 Halobacte
14	68.8	8.8	4262	1	AF335723	AF335723 Burkholde
15	64.4	8.2	14483	1	AE006992	AE006992 Mycobacte
16	64.4	8.2	67200	1	MTV017	AL021897 Mycobacte
17	57.6	7.3	1983	6	AX047403	AX047403 Sequence
18	55	7.0	36368	1	SC9B5	AL035206 Streptomy
19	54.8	7.0	346897	1	AP002995	AP002995 Mesorhizo
20	54.6	6.9	45313	1	SCD95A	AL357432 Streptomy
21	54.6	6.9	125020	9	AF429315	AF429315 Homo sapi
22	54.4	6.9	38995	1	SCF34	AL109974 Streptomy
23	54	6.9	34182	1	SC111	AL096849 Streptomy
24	53.6	6.8	77534	1	AF235504	AF235504 Streptomy
25	53	6.7	33779	1	SCG03	AL096882 Streptomy
26	52.8	6.7	125020	9	AF429315	AF429315 Homo sapi
27	52.2	6.6	1849	1	SGSTRB1F	X78973 S.galbus DS
28	52	6.6	349116	1	AP003003	AP003003 Mesorhizo
29	51.4	6.5	3644	5	AF034576	AF034576 Gallus ga
30	51.2	6.5	22115	1	SCC22	AL096839 Streptomy
31	50.6	6.4	5525	1	SGSTRB1FG	X78974 S.glaucisce
32	50.6	6.4	14596	1	AE007168	AE007168 Mycobacte
33	50.6	6.4	25459	1	SGAP6985	AJ006985 Streptomy
34	50.6	6.4	30302	1	MTCY6G11	Z92774 Mycobacteri
35	50.6	6.4	40806	1	MSGY23	AD000016 Mycobacte
36	50	6.4	29625	1	SC422A	AL159178 Streptomy
37	50	6.4	36394	1	SCD35	AL160312 Streptomy
38	50	6.4	40104	1	SCF11	AL132662 Streptomy
39	49.8	6.3	36734	1	SCG610	AL049497 Streptomy
40	49.6	6.3	26555	1	SC261	AL331014 Streptomy
41	49.6	6.3	146555	2	OSJN00013	AL606448 Oryza sat
42	49.6	6.3	215050	1	AL646057	AL646057 Ralstonia
43	49.4	6.3	150814	2	AP003541	AP003541 Oryza sat
44	49.2	6.3	12130	1	AE005133	AE005133 Halobacte
45	49.2	6.3	41173	1	SC141	AL132648 Streptomy

## ALIGNMENTS

RESULT 1	AX339682	786 bp	DNA	linear	PAT 10-JAN-2002
LOCUS	AX339682	Sequence 3 from Patent WO0196382.			
DEFINITION	AX339682				
ACCESSION	AX339682				
VERSION	AX339682.1	GI:18135684			

### KEYWORDS

ORGANISM Rhodothermus marinus.  
Rhodothermus marinus

REFERENCE 1 (sites) Bacteria: CFB group; Rhodothermus group; Rhodothermus.  
Wichner, K.B., Holst, O.P., Hachem, M.Y., Karlsson, E.M. and Hreggvidsson, G.O.

TITLE Thermostable cellulase  
JOURNAL Patent: WO 0196382-A 3 20-DEC-2001;

FEATURES Prokaria enf. (IS)

location/Qualifiers

1..786

/organism="Rhodothermus marinus"

BASE COUNT 141 a 236 c 279 g 130 t

ORIGIN





```
FEATURES
    source      Netherlands
                Location/Qualifiers
                1..1116
                /organism="Streptomyces sp. 11A68"
                /strain="11A68"
                /db_xref="taxon:133452"
    gene
                1..1116
                /gene="cell12A"
    misc_feature
                1..93
                /gene="cell12A"
                /note="Region: signal sequence"
    CDS
                1..1116
                /gene="cell12A"
                /note="endocellulase"
                /codon_start=1
                /transl_table=11
                /product="cellulase 12A"
                /protein_id="AAF91283.1"
                /db_xref="GI:9651813"
                /translation="MNSHRSATMTVLVVLASGLALTAAPADAQNOICDRYGTITI
                ODHYVONNRNCTSATOCINVTNGSEFETQADGSPYTNCAKPSYSPYDGHYGNCA
                PRTLLPMRISISGASPSVSYRYTNGVYNAADIMLDPRTNGVYRTEIMIPNFG
                PVQPIGSPVGTAGHVGSRWETWGSNDVLSPLASPISSMSFPEKDEVOAASHG
                LATPMWLTSLIAGREPMRGCTGLAVNSPSSAVNMGCGMGTGCPAPAAOVSYRHTY
                PCGFTVDTITNTGTSPVDGMEIDLTLPAGHTVTSWMAALISPASAVTARSTGSNGR
                IANGTQSFQGTSSGAGTAPAGARLNGTSCIVR"
                94..756
                /gene="cell12A"
                /note="Region: catalytic domain"
                757..792
                /gene="cell12A"
                /note="Region: linker region"
                793..1113
                /gene="cell12A"
                /note="Region: cellulose binding domain"
BASE COUNT    191 a 432 c 340 g 153 t
ORIGIN
Query Match      16.8%; Score 131.8; DB 1; Length 1116;
Best Local Similarity 50.7%; Pred. No. 3.4e-10;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

QY 98 ctgaagcccgagccgacgctgagctgtgcgagcgtggagcgcgcgcgatgtggccggg 157
    |||||
DB 77 CGGCTCCCGCCAGGGAACCGACAGATCTGCGACCGCTACGGCACCACCGATCCAGG 136
QY 158 ggcgtacacgggtgatcaacaacglatggtggcgcgagagcccgatgtcatggaatcg 217
    |||||
DB 137 ACCGGTACGtGtGtGCAACACCGCTGGGGCACACCGCACCGCATGATCATATGTA 196
QY 218 gactggaacgggcaacttcaacatcacacggtcgatcagaacaacggaacacgttg 277
    |||||
DB 197 CCGGCAACGGTTTCGAGATCACCCAGGCGACGTTTCGTCGCGACCAACAGGCGCCCGA 256
QY 278 ccgcatatccggcatctacttctgggtgacactggggcgcgctgcaagagcaattcgat 337
    |||||
DB 257 AGTCTATTCCTCGTCTACGAGCGCTGCCACTACGGCAACTGCGGCCGCCGACGACGCG 316
QY 338 tgcgcgagcgctgtgcaggaactgtccgaagtgcgcaagagctgagcgctacgcccga 397
    |||||
DB 317 TGCCCATGCGGATTCAGCTGCATCGGCGAGCGCGCCGACGACATGTCTCTACCGTACACG 376
QY 398 cgaagggcgctgtgaaatgcgcctacgacatctgtgtcaagtcaccgtcaagaattccgga 457
    |||||
DB 377 GCAACGGCGTCTACACAGCGCGGTACGACATCTGG-----CTGGACCGCACACCCGCA 430
QY 458 acgggtacagcgcgcgcgcgatgtatgtctggtcgtgaactgtgaaagcgcggtgtatgc 517
    |||||
DB 431 CCAACGGGGTGAACCGGACCGAGATCATGTCTGTTCAACCGGGTGGCGCCGTCACGC 490
QY 518 cggcgcgagcgcgctgtgacacgctgtgaaatgtgcggggcgcaactgtggaagctgtgatg 577
    |||||
```

```
DB 491 CCATCGGTTCCGCGGTGCGCACGGCCCACTCGCGCGCCGACAGCTGGAGAGTGTGACCG 550
QY 578 ccgactgggactgtgaattacatcgcctaccgagcgcaagcagccacacgactcggtgagcg 637
    |||||
DB 551 GCAGCAACGGTTGCAACGACGATGATCTCTCTGCGCGCCCTCGCGGATCAGACGCTGGA 610
QY 638 agctggaacctgaagccttcacatcagacgagcggtgcgcggcggtcatcatccggccgag 697
    |||||
DB 611 GCTTCGACGCTCAAGGACTCTGTCGACCAAGCGCTGACGACCGGCTCGGCGCGGACT 670
QY 698 gttatctgcatgtggtgtgagagcgaggtctgaaactctgggagagcgggcggtctgcgaa 757
    |||||
DB 671 GGTACCTCACACCACTCCAGCGGCGCTTCGAGCGGCGGAGGCGGCGGACCGGCTGCGCG 730
QY 758 ggcgcgattttccgtaacggtgca 782
    |||||
DB 731 TGAACCTGTTCTCTCCGCGGTGA 755

RESULT 5
ARI29926 1116 bp DNA linear PAT 16-MAY-2001
LOCUS ARI29926
DEFINITION Sequence 2 from patent US 6187577.
ACCESSION ARI29926
VERSION ARI29926.1 GI:14117823
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1116)
AUTHORS Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Weyler,W.
TITLE Cellulase producing Actinomyces cellulase produced therefrom and
method of producing same
JOURNAL Patent: US 6187577-A 2 13-FEB-2001;
FEATURES
    source      Location/Qualifiers
                1..1116
                /organism="unknown"
BASE COUNT    191 a 432 c 340 g 153 t
ORIGIN
Query Match      16.8%; Score 131.8; DB 6; Length 1116;
Best Local Similarity 50.7%; Pred. No. 3.4e-10;
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

QY 98 ctgaagcccgagccgacgctgagctgtgcgagcgtggagcgcgcgcgatgtggccggg 157
    |||||
DB 77 CGGCTCCCGCCAGGGAACCGACAGATCTGCGACCGCTACGGCACCACCGATCCAGG 136
QY 158 ggcgtacacgggtgatcaacaacglatggtggcgcgagagcccgatgtcatggaatcg 217
    |||||
DB 137 ACCGGTACGtGtGtGCAACACCGCTGGGGCACACCGCACCGCATGATCATATGTA 196
QY 218 gactggaacgggcaacttcaacatcacacggtcgatcagaacaacggaacacgttg 277
    |||||
DB 197 CCGGCAACGGTTTCGAGATCACCCAGGCGACGTTTCGTCGCGACCAACAGGCGCCCGA 256
QY 278 ccgcatatccggcatctacttctgggtgacactggggcgcgctgcaagagcaattcgat 337
    |||||
DB 257 AGTCTATTCCTCGTCTACGAGCGCTGCCACTACGGCAACTGCGGCCGCCGACGACGCG 316
QY 338 tgcgcgagcgctgtgcaggaactgtccgaagtgcgcaagagctgagcgctacgcccga 397
    |||||
DB 317 TGCCCATGCGGATTCAGCTGCATCGGCGAGCGCGCCGACGACATGTCTCTACCGTACACG 376
QY 398 cgaagggcgctgtgaaatgcgcctacgacatctgtgtcaagtcaccgtcaagaattccgga 457
    |||||
DB 377 GCAACGGCGTCTACACAGCGCGGTACGACATCTGG-----CTGGACCGCACACCCGCA 430
QY 458 acgggtacagcgcgcgcgcgatgtatgtctggtcgtgaactgtgaaagcgcggtgtatgc 517
    |||||
DB 431 CCAACGGGGTGAACCGGACCGAGATCATGTCTGTTCAACCGGGTGGCGCCGTCACGC 490
```



QY	518	cggggagcagcgcggtgagccacgcttgagacgcccgggcccacctcgggaagtctgtatg	577
Db	491	CCATGGGTTCCCGCGTCGGCACGGCCACGTGGGGGGCCGACGCTGGGAGGTGGACCG	550
QY	578	ccgaactggagacttggaaattacatcgccctacccggcgcgacgagcccaaccgltgtgtacg	637
Db	551	GCAGACACGGTTTCGAAACAGCGATCTCTTCCTGGCGCCCTCGCGCATTCAGACATCGGA	610
QY	638	agcttggaccttgaaggccttcatacgacgaacgcggtgtgccggcgatacatccggccggagt	697
Db	611	GCTTCGACGTTAAGGACTTCGTGCACACGGCGTGTAGCAGCAGGCGCTGTGCACCCCGGACT	670
QY	698	gatactcgcacgtcgggtggaagcgggcttcgaaactctggagggcgggcgctctcgaa	757
Db	671	GGTACCTACACGAGATCCAGCGGGGCTTCGACCCTGGGAGGGCGGACACGGTCTGGCGG	730
QY	758	gcgcgcatllttccglaacggtgca	782
Db	731	TGAATCTGTTCTCTCCGCGGGTGAA	755

[illegible]

Db	431	CCACGGGGGTACCGGACCGGACGATCATGATCTGGTTACCGGGTCGGCCGGGTCCACG	490
OY	518	CGGCGCGCAGCCGCGCTGCGCCACCGTGTGAACtgcgagcgccacctcgtggaagtcgtatg	577
Db	491	CCATCGGTTTCGGCTCGGACGAGGCCACGTCGGCGCGCCGAGCTGGGAGGTGTGGACCG	550
OY	578	ccgacttgcgacttgcgattacatcgcctcaccgcgcaagcagccacacacgttcgctgcg	637
Db	551	GCAGCAGCGTTTCGACGACCGTGATCTCTTCTCTGGCCCTCCCGCATCAGCAGCTGGA	610
OY	638	agctgcgaccctgaagccttcacacgaagacgcggtctgcgcggtacatccgcgcgagtc	697
Db	611	GCTTCGAGCGTCAAGGACTTCGTGCAGCAGAGCCGTGAGCAGCAGCCCTGGCCACCCCGGACT	670
OY	698	ggtatctgcatacgcgctgcgagacgagcttcgaactcttcggagggcgcggttcgcgaa	757
Db	671	GGTACTTCACCGACGATCCACGGCGGCGCTTCGAGCCTGGGAGGGCGGCAACCGTCTGGCGG	730
OY	758	gcgcgcgatttttcgtaacgctgca	782
Db	731	TGAACCTCGTTCCTCCGCGGGTGAA	755

[illegible]

QY	641	tgagactgaagccctcatcga-gcaagcagatcgccgcgcgtacatccgcccagatcg	699
Db	1065	TCGAGCTGAAGGACTTTCGTGCACCAAGCCGCTACAGCAGGCGCTTGCCACCCCGAGCTGG	1124
QY	700	tatctgcatctcggctggagacagcgcttcgaactctggagggcgggcgccgtctcgaaagc	759
Db	1125	TACCTCACACAGCATCAGGCGGGCTTCGAGCGGTGGAGGGCGGACCGGCTTGCCCGTGG	1184
QY	760	gccgattcttcctgaagcgtcga	782
Db	1185	AACTGCTTCCTCCTCCGCGGTGA	1207
RESULT	8		
SHU51222			
LOCUS	SHU51222	3527 bp	DNA linear BCT 23-UN9-1997
DEFINITION	Streptomyces halstedii beta-1,4-endoglucanase (celA2) gene, complete cds, and cellulose binding protein (p40) gene, complete cds.		
ACCESSION	U51222		
VERSION	U51222.1	GI:2209259	
KEYWORDS			
SOURCE	Streptomyces halstedii.		
ORGANISM	Streptomyces halstedii.		
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
AUTHORS	1 (bases 1 to 3527) Garda, A.L., Fernandez-Abalos, J.M., Sanchez, P., Ruiz-Arribas, A. and Santamaría, R.I.		
TITLE	Two genes encoding an endoglucanase and a cellulose-binding protein are clustered and co-regulated by a TTA codon in Streptomyces halstedii JMB		
JOURNAL	Biochem. J. 324 (Pt 2), 403-411 (1997)		
MEDLINE	97307849		
REFERENCE	2 (bases 1 to 3527)		
AUTHORS	Santamaría-Sánchez, R.I.		
TITLE	Submitted (13-MAR-1996) Ramon I. Santamaría-Sánchez, Instituto de Microbiología Bioquímica, Consejo Superior de Investigaciones Científicas / Universidad de Salamanca, Avda. del Campo Charro, s/n, Salamanca, Salamanca 37007, Spain		
JOURNAL	Location/Qualifiers		
FEATURES	1..3527		
source	/organism="Streptomyces halstedii"		
gene	/strain="JMB"		
gene	/db_xref="taxon:1944"		
gene	595..1728		
gene	/gene="celA2"		
gene	595..1728		
gene	/gene="celA2"		
gene	/EC_number="3.2.1.4"		
gene	/note="beta-1,4-endoglucanase"		
gene	/codon_start=1		
gene	/transl_table=1		
gene	/product="cel2"		
gene	/protein_id="AAC5429.1"		
gene	/db_xref="GI:2209260"		
gene	/translation="MRALPHGPRTPRSLGALLALATVAALAAAPAAHADLTVCCEQYGSSTIGRIVYVONNRWGASAPCVATDSEFRYTOADGAVPTNGAPKSPYVNGCGHVTNSPGNLPAQVYSGASAPSSISYGVGAAVYNASYDIDLPTPKNGVNRREIMIMWLNKVGPIOPTGSOAGTASVSGRTWQVWRGSGNDVTSFYAPBAVASMSFTDMDPYRANLIRGAAQNNWYLLTSVQAGEPMONAGLAVNSFTVNLGTGTSFGAPGAPVTRACVAVATNWSMDGFTADVTVANSTGSDVDSRLSFTLPSGGRVNMANMATVSGPTGAVATATGLASHAQVAAAGASQTFGQGTSGFSPKSGFSINGARCA"		
gene	1945..3039		
gene	/note="cellulose binding protein"		
gene	/codon_start=1		
gene	/transl_table=1		
gene	/product="p40"		
gene	/protein_id="AAC5430.1"		
gene	/db_xref="GI:2209261"		
gene	/translation="MARKRRLVSLTVAFATLLGGLALLTLFGGNAQAHGVYTMTPGSRFTYGLWLAKSTGSLDPTNPAKKAALSESQNALYNNFAVDNSNAGRGADYVYDGKIL		

BASE COUNT 632 a 1270 c 1106 g 528 t  
ORIGIN  
CSACDASRPYNPTGNAARSDMPRTHTAGARTIQKHSNNMAAPSCSFVYLTSKPGYSPS  
TELADMDLELEIETVNPOTGSPCTDGGHYWMLDPSRGGDVAMFLOWVRSSOEN  
FFGCSADVDFDGGNGEAMGIRBSGSGTPTPTPTPTPTPTPTDPSGGMAAYRYNVMYS  
CGGCGSEVMNHSSTTARADGAAWKMTPCAGAAVSVMNAGALTTSGDCAVTVRSLDYNNS  
IPDPSGVTPTGTTATGTCNLPVSGICVNP\*

Query Match	Score	DB 1	Length
12.48	97.8	DB 1	3527

Matches	304;	Conservative	0;	Mismatches	317;	Indels	6;	Gaps	1.
---------	------	--------------	----	------------	------	--------	----	------	----

QY 156 ggagcgctaccggtgatcaacaacgtatggcgcgagaccgccagtgcattgagtc 215

Db 747 GGGCCGCTATGTGGTCCAGAACACCGCTGGGGCGGAGTGCCCCCAGTGCGTCACCGC 806

QY 216 cggacttgaaacgycacttcacgatcacacggccgatcacgacacgcaacaagt 275

Db 807 TACCGACGCGCTTCGGGTGACCCAGGCCGACGGCGCCGTCCCCACGACGGTGCCCC 866

QY 276 ggccgctatccggccatctactcgggtgccactggggcgccctgcagagcaattcggg 335

Db 867 GAAGTCCCTATCCCTCGGCTCTCAACGGCTGTCACTACACCACTGCTGCGCCGGGACGAA 926

336 attgacgcgcgtgcagagctgtccgacgtgcgcagcgtggaacgtcacgcgat 395

Db 927 CCTCCCGGCGAGGTATCGCCAGCGCCCGCAGCATCTCCTACGGCTTGT 986

QY 396 caccgacggccgctgtaatgcgcctacgacatctgttcagtcgccgtcagcaaatccgg 455

Db 987 CGGCAGCGCCGTGTACCAAGCGTCTGGCTGGACCCACACCCAGAAGAA 1046

QY 456 caacgctacgagcgccgagctgtatctgtgctgaactggaacggcggcgtgat 515

Db 1047 CCGGGTGAACC-----GTACCGAGATCATGTGCTCGGCTCAACAAGGTCGGCCCGATCCA 1100

QY 516 gccgagcgcagccgctggccaaccgtggaagtctgtca 575

Db 1101 GCCCATCGGCTCGCAGCGCCGCTCCGTCGGCGGGCGCACCTGGCAGGTGTGGAG 1160

QY 576 tgcgactggaattacatcgccctaccgagcagacgcccacacgctcggtgag 635

Db 1161 GGGCAGCAACGGCTCCACAGACGTATCTCCTTCGTGCGCCCTTCGGCCGTGCGCCAGCTG 1220

QY 636 cgaagctggaaccttgaaagcccttcatacgaaacgacgctgcgccccgcgcgtatacatccgcgcgcga 695

Db 1221 GAGCTTCGACGTCATGGACTTCGTCGCGGACACCATCGCCCCGGGCA TGCGCGAGACAA 1280

QY 696 gtgttatctgcatagcgttgagacgggcttcgaactctggagggcggggccggtcttcg 755

Db 1281 CTGGTATCTCACCAGTGTCCAGGCGAGGATTCGAGCCGTGGCAGAACGGTGCCTGGACTCGC 1340

QY 756 aagcgccgatttttcggtaacggtgca 782

Db 1341 GGTGAACCTCCTTCTCCTCAACCGTGA 1367

## RESULT 9

SREGLS			
LOCUS	1470 bp	DNA	linear
SPECTS			RCT 16-DEC-

DEFINITION	S.rochei egls gene.
ACCESSION	X73953

VERSION	X73953.1	GI:393391
KEYWORDS	cellulase; egls gene; endoglucanase	

SOURCE Streptomyces rochei.  
ORGANISM streptomycetes rochei

**Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyc**

REFERENCE	1 (bases 1 to 1470)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	
101	
102	
103	
104	
105	
106	
107	
108	
109	
110	
111	
112	
113	
114	
115	
116	
117	
118	
119	
120	
121	
122	
123	
124	
125	
126	
127	
128	
129	
130	
131	
132	
133	
134	
135	
136	
137	
138	
139	
140	
141	
142	
143	
144	
145	
146	
147	
148	
149	
150	
151	
152	
153	
154	
155	
156	
157	
158	
159	
160	
161	
162	
163	
164	
165	
166	
167	
168	
169	
170	
171	
172	
173	
174	
175	
176	
177	
178	
179	
180	
181	
182	
183	
184	
185	
186	
187	
188	
189	
190	
191	
192	
193	
194	
195	
196	
197	
198	

TITLE	Direct Submission
JOURNAL	Submitted (07-JUL-1993) C. Mastromei, Università di Firenze, Dipartimento di Fisica







```

mlsc_feature
1822..2373
/gene="SCG11A_02"
/note="Pfam match to entry PF00480 ROK, ROK family, score 131.80, E-value 4e-38"
3041..3646
/gene="SCG11A_03"
3041..3646
/gene="SCG11A_03"
/note="SCG11A_03", probable amidase, len: 201 aa; similar to SW:AMP_ECOLI (EMBL:X15237) Escherichia coli anhydro-N-acetylmuramyl-L-uridipside amidase Ampd, 183aa; fasta scores: opt: 211 z-score: 267.1 E(): 1.7e-07, 39.5% identity in 114 aa overlap and to Streptomyces coelicolor SC2A11.21c; fasta scores: opt: 674 z-score: 751.6 E(): 0; 50.0% identity in 202 aa overlap. Contains match to Pfam entry PF01510 Amidase_2, N-acetylmuramyl-L-alanine amidase"
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="CAB61584.1"
/db_xref="GI:6468424"
/translation="MERETAPVPTRRRLTGALATVPYTLTSGTAAACVRAVDYPSAEWLPASTNSRSSRSTAPYVDFVGVVWVHTQETVADLTIFRDPKQVSAHYVRSRSGHVAQCVRETDIAHNAQNDYINFRKISIEHEGWDRPETYTNMEQSAALYACTT3191..3637
/gene="SCG11A_03"
/note="Pfam match to entry PF01510 Amidase_2, N-acetylmuramyl-L-alanine amidase, score 89.80, E-value 5.7e-23"
3811..3815
3821..5080
/gene="SCG11A_04"
3821..5080
/gene="SCG11A_04"
/note="SCG11A_04", possible transcriptional regulator, len: aa; similar to TR:OJ1551 (EMBL:Z99108) Bacillus subtilis transcriptional regulator AcOR, 605 aa; fasta scores: opt: 393 z-score: 446.5 E(): 1.7e-17; 34.7% identity in 199 aa overlap"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB61585.1"
/db_xref="GI:6468425"
/translation="MTDAWLALPGCAQDVEERARALRHHETFEAGTVRRPVAVVADSMRSVRVAGVDPDGTASVELMDGDLGATRAHNPISRYKVPYRLTGLTFPADGELLVACDARLILWEGHPATRRRAGRMNFVPEARSESAYGVGNAGTAVAGRPVQVAAAEHFI1RRQVPTMCAAPVNDRTGVLGADVDTISGGLIAPHNSGFVQAVARAEEQSLALLTPTERASAEALTEVLSGRDEALISADGRKRVLSRSHSEITVLVLAHNPGLTQDELICALLYEDVEVPVTLTAAELATVLRGLIGPRLASRPYRLTMEVSEPTSVERRLRGAVTAAALAAVAPRLGSOAPRVGRLRRRLADGRLAALILACGDPDLADMAHNAFWGDEDDLDVWRA1LAAYVPTAATSSRLAALSELAAIRPW"
5205..5209
5218..6741
/gene="thca"
5218..6741
/gene="thca"
/note="SCG11A_05, thcA, aldehyde dehydrogenase, len:554 aa; highly similar to SW:THCA_RHOSN (EMBL:U017129)

11.1%; Score 87.6; DB 1; Length 41782;

```

Best Local Similarity	47.5%	Pred. No. 0.00033:	Matches 298:	Conservative 0:	Mismatches 324:	Indels 6:	Gaps 1:
Qy 156	gggagcgctacccggytgatcaacaacgtatgggagcggaagacgcgcacgtgcatgaggt	215					
Db 27838	ggcgaggtacgtctgcgcagaaacacccgtgggggtccacccaccacccgctgcacacgcg	27887					
Qy 216	cggactgtgaacacggcgaacttcagcatcaacacggtgcgcatcaacgaacacggcaacaagt	275					
Db 27888	caccgacacacccgcttcgggtacacccagccgacggctgcgcacccagccaaacggggccc	27947					
Qy 276	ggccgctcatcccgccacatctacttcggtgtgcacatggggcgctgtcaacgaacatcggg	335					
Db 27948	gaagtcgtaaccctggtggtctcaacggctgcacacacgaacattgttcacccgggacggcg	28007					
Qy 336	attgcgcggggcgctgtgcagagagctgtctcgaagttggcagcgaggtctgagaccccaacgcat	395					
Db 28008	cctcccccctccggctgcacacccgtctccggcgccgctgcacacatctctgacgcttgcgt	28067					
Qy 396	caacgagcgccgctgtgaatgtccgcctcaacacatctgtgtacgtccctgcacgaattccgg	455					
Db 28068	cgacggcgcccgctctcaacacccgtctgacacatgag-----ctggacacccgacggcccg	28121					
Qy 456	caacggtacaacgacggcgacgcgcgagctgtatctgtcgtgaactgtgaaacggcgctgat	515					
Db 28122	caccacacgggggtgaacacacgacgaatattatctgtgtaaacagcgtgggtccgatcca	28181					
Qy 516	gccggcgcgacgcgcgcgtgtgcacacgctgtgaactgtgcggcgacacgtgtggaatctgt	575					
Db 28182	gcccatcgcgcttaccgggtgggacacggcctccgctcgccggcgacacatctggaggtctgag	28241					
Qy 576	tgccgaactgtggaactgtgaattacatctgcctaccgcgcacgaacgccacacgctggtgag	635					
Db 28242	cgccggcgaacgctctgcagacacgctcctctgctgtctgtgcacacgctgcggaacacggcgtg	28301					
Qy 636	cgagctgtgacacctgaagagcctcatctgcagaacgcggtgtccgcgcgcatactccgcgcga	695					
Db 28302	gagcttgcacactgcatagacacttgcgtccggggcaccgctccgcgcgcacactcgcgcgaacaga	28361					
Qy 656	gtgtgatactgcatacgcgtgtgtagaacgggctctggaactctggagggcgggcgctgtcg	755					
Db 28362	ctggactactgcacgacacgcttcagcggcggtgttcgacccgtgcagaaacggcgccgactgtgc	28421					
Qy 756	aagcgcgcatctttccgcgaacggtgcag	783					
Db 28422	cgtgacactcttctctccaccgtgcag	28449					
RESULT 13	AE005123	11202 bp	DNA	Linear	BCT 12-FEB-2001		
LOCUS	AE005123						
DEFINITION	Halobacterium sp. NRC-1 section 154 of 170 of the complete genome.						
ACCESSION	AE005123						
VERSION	AE005123.1						
KEYWORDS	GI:10581826						
SOURCE	Halobacterium sp. NRC-1.						
ORGANISM	Halobacterium sp. NRC-1.						
	Archaea; Euryarchaeota; Halobacteria; Halobacteriales;						
	Halobacteriaceae; Halobacterium.						
REFERENCE	1 (bases 1 to 11202)						
AUTHORS	Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M., Shukla,H.D., Lasky,S.R., Balliga,N., Thorsson,V., Sbrogna,J., Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Weiti,R., Gao,Y.A., Leitnauer,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W., Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H., Isenberger,T.A., Peck,R.F., Pohlischrod,M., Spidlich,J.L., Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P., Omer,A.R.D., Ebdardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and Dasarma,S.						
TITLE	From the cover: genome sequence of halobacterium species NRC-1						
JOURNAL	Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)						
PUBMED	11016950						
REFERENCE	2 (bases 1 to 11202)						

Ng, W. V., Kennedy, S. P., Mahatas, G. G., Bernstein, B., Pan, M., Shukla, H. D., Lasky, S. R., Baliga, N., Thorsson, Y., Shrogha, J., Swartzell, S., Weir, D., Hall, J., Dahn, T. A., Weller, R., Goo, Y. A., Leitlhauser, B., Keller, K., Cruz, R., Danson, M. J., Hough, D. W., Maddocks, D. G., Jablonksi, P. E., Krebs, M. P., Angele, C. M., Dale, H., Jansenberger, T. A., Peck, R. F., Pohlschod, M., Spädic, J. L., Jung, K. H., Alam, M., Freitas, T., Hou, S., Daniels, C. J., Dennis, P. P., Omer, A. D., Eberhardt, H., Lowe, T. M., Liang, P., Riley, M., Hood, L., and Casarrema, S.

Submitted (14-JUL-2000) Institute for Systems Biology, 4225  
Roosevelt Way NE, Seattle, WA 98105, USA

**Location/Qualifiers**

```

/gene="meta"
/ncbi="meta"
/codon_start=1
/transl_table=1
/product="homoserine O-acetyltransferase"
/protein_id="AAG20507.1"
/db_xref="GI:10581827"

```

```
CDS
        /gene="hal"
        /note="VNG2421G"
        1351..2646
        /gene="hal"
        /note="Hal"
        /codon_start=1
        /transl_table=11
        /product="O-acetyl homoserine"
        /protein_id="AAG20508.1"
        /db_xref="GI:10581828"
```

```

CDS
      /gene="g1cd"
      /note="ANG24.22c"
      2715..5753
      /gene="g1cd"
      /note="g1cd"
      /codon_start=1
      /transl_table=1
      /product="glycolate oxidase subunit"
      /protein_id="Ang20509.1"
      /db_xref="GI:10581829"

```

CDS

CDS

CDS

```
/product="D-Lactate dehydrogenase"  
/protein_id="AAG20513.1"  
/db_xref="GI:10581833"
```

/translation="M.SALVADHIIHTTKNGFAFTTLMGLVYALLIOPSGRSLDLSRK  
 TLIVTIGLMASSGPFVYVATARSVALIIVGAIVYGTGGLPTARALVADLYSKQR  
 AFCHHTASGDLGMSAAGIAVALALLSMRIARVPIYICLTAIAVALAHWSRESYFE  
 RRSILAVDTARLLSTQFRRWILVAVLYAFWQWAIAPLPTFTDSEKRAPVAVGYVA  
 FGSLAEVAGAIARPPACTGLSDTISRRAVALVALYGAALSAVVVPSPVAVGAVVVF  
 AAGLAFPPVWQSYLMDAPFPESAGDGLGAMRTVYIGLGLPTVYGAVSTAANYEIA  
 FVGLVGLIYAVIIVART"  
 complement(9890..11137)  
 /gene="thrc1"  
 /note="VNG2430G"  
 complement(9890..11137)  
 /gene="thrc1"  
 /note="thrc1"  
 /codon\_start=1  
 /transl\_table=1  
 /product="threonine synthase"  
 /protein\_id="AAG20514.1"  
 /db\_xref="GI:10581834"  
 /translation="MTDLSLAGETPPACATDGVWLAVCECGTLPFADVVYRGCEPSG  
 LLEKRYADYPGFEAFSGRWVRYADALPVVAGVSGVSGDTPLYEAPOVAAAAGVADVR  
 VKHESRPTGSGDRGMTVGRVARELVGVRLACASGNTSAAALAAVGAADPVVILV  
 LPACKVAGRIAOALAGARICEVDNGFDDCLDVAFRLARGEAYILNLSINPFLBEO  
 KTIKFEILBOSRTATGEMPRIVLPVNMAGNTALLYAFRELVAAAGMTDDEMPALV  
 VQAGAAPVNEAVRAGSKRTERMTDVEFRATARITGNPVNAPKALPEVYFTGGTAALV  
 SDDITAOQRMLAGDGIIVGEPSAASVAGLLALEREETISADERVCLTTGHLIKDPA  
 AAAAAGGDTTVPVPADAGVLDAL"  
 1670 a 4414 c 3626 g 1492 t  
 BASE COUNT  
 ORIGIN

Query Match	8.8%;	Score 69;	DB 1;	Length 11202;
Best Local Similarity	43.8%;	Pred. No. 0.23;		
Matches 300;	Conservative 0;	Mismatches 385;	Indels 0;	Gaps 0;

OY	69	cgatgagcaacaaaggaaagaaacccggaacctgagagccgaagccgacccgagctgtgtg	128
Db	7058	CGACCCCTTACATCGGCGAGGACCGGGGCGACAGCTTCGGCGGAGCGTGTGCATTCGA	7117
OY	129	acgcctggaagcgcgcgcgaatgtgtgcgcgggggcgcctaccgggtgtatcaacaacgtctgg	188
Db	7118	GACCTGCGTCGGCGCGCGGGAATTCCTCAGCATCCACGTCCTCCACTCCACGACGAGAACCGA	7117
OY	189	cgcggaagaccccgatgtcatgtgaagcttggaaacgycgaacttcacgatcacg	248
Db	7178	CGGCTTCATCGGCGGAGCGCCGAATCTGCCCGGATGAGCGGGCTACGTATCAACGTCGC	7237
OY	249	ggcgcgaatcaagcaaacgcgcaacaagctgcgcgcctatccgcacatctcatcttcggtgca	308
Db	7238	CGGCGGCGCGCTGTCTGACGAGAGCGCCTGTGCGCACGCGCCCGCAGGAGCGGCTGATTCG	7297
OY	309	ctgggagcgcctgcaagagcaatctcggaatgtgcgcgggcgtgtgcaaggagcgtgccagct	368
Db	7298	GGGGGCGCCCTCGACGCTTCCGACCGAGACCGTCCCGCGGCGTCCCGCTCCACGA	7357
OY	369	gcgcagagcctggagcgtctcaagccgaatcaagacgycgcgtggaaatctgcgcttcagacat	428
Db	7358	CGCCACGACATCATCACACGACCGCCACACTGCGGCGCTCCACGAAGCGCGCCCGCAGAGAA	7417
OY	429	ctggtctcagctccgttcaagaattccggcaagcgtctacagcgcgcgcgcgcgcgcgtatg	488
Db	7418	CTCTCCACCGACACACCGCGCACAGCTGTGTGCGCGCTCCCGCGCAGCCCGCTGTGAA	7477
OY	489	ctggagctgaagctggaacgcgcgtgtatgtgcgcgggcgcgcgcgcgcgcgcgcgcgcgaact	548
Db	7478	CGCGCTGAACCGCCCGCAGCGCGCGGAGCGGTCCGCGTTCACCGCATCCGCCCGTTCGTGA	7537
OY	549	ggcgcgggcaccccttgggaagctctgtatctgcgcgaacttggagctgaatcatctgcgtccacg	608
Db	7538	CCTCGCGAGAACCGCGGACACGTCGCGCGGCGCGCTGTTCGACCCCGCGCATCGAGCGGT	7597
OY	609	gcgcacagcgcgcacacacgctgcgtgtagcgcgcgtgaccttgaagcctcatctcgaagcgc	668
Db	7598	CGACCTCATGTAACGAGCGGAGGTCGCGGAGAGACCTCCACTGCTACGAGCGGCGCG	7657

[illegible]

RESULT	14				
AF335723					
LOCUS		4262 bp	DNA	linear	BCT_11-FEB-2001
DEFINITION	Burkholderia pseudomallei genome hydrogenylase subunit 4 and hydrogenase-4 component B forms, complete cds.				
ACCESSION	AF335723				
VERSION	AF335723.1	GI:12744948			

**SOURCE**  
**ORGANISM**

## REFERENCE

TYPE

## REFERENCES

JOURNAL

Source

CDC

ORIGIN

Query Match	8.8%;	Score 68.8;	DB 1;	Length 4262;
Best Local Similarity	47.3%;	Pred. No. 0.31;		







---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 13:37:08 ; Search time 544 Seconds  
(without alignments)  
2480.689 Million cell updates/sec

Title: us-10-003-759-3

Perfect score: 786  
Sequence: 1 atgaacgtcatgctgcggt.....tttcgcgaacgtgtag 786

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: N\_Geneseq\_032802:\*  
2: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
3: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
4: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
5: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
6: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
7: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
8: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
9: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
10: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
11: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
12: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
13: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
14: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
15: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
16: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
17: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
18: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
19: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
20: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
21: /SID5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
22: /SID5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
23: /SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
24: /SID5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	131.8	16.8	1116	20 AAV72359
2	131.8	16.8	1116	21 AA257029
3	131.6	16.7	1715	21 AA257031
4	57.6	7.3	1983	22 AAA54513
5	53.6	6.8	77536	21 AA14651
6	49.8	6.3	756	17 AA16766
7	49.4	6.3	1404	21 AAA38763
8	49	6.2	756	14 AA043032
9	48.8	6.2	985	19 AAV64548

10	48.8	6.2	985	19 AAV44439
11	48.8	6.2	985	20 AA219349
12	48.8	6.2	985	20 AA219137
13	48.4	6.2	12152	22 AA508699
14	48	6.1	4403765	22 AA199683
15	47.8	6.1	24379	18 AA793095
16	47.8	6.1	24379	18 AAV25925
17	47.6	6.1	30001	19 AAT61016
18	47.6	6.1	30001	19 AA05110
19	47.2	6.0	1908	16 AA099364
20	47.2	6.0	1908	16 AAV84065
21	47.2	6.0	1908	21 AAC61403
22	46.8	6.0	15872	21 AA287283
23	46.8	6.0	20394	22 AAF24892
24	46.6	5.9	2712	19 AAT94214
25	46	5.9	44377	18 AAT78508
26	46	5.9	44377	18 AAT80414
27	45.6	5.8	744	18 AAT74073
28	45.2	5.8	1291	20 AA87940
29	45.2	5.8	2852	19 AAV64558
30	45.2	5.8	2852	19 AAV44449
31	45.2	5.8	2852	20 AA219359
32	45.2	5.8	2852	20 AA219147
33	45.2	5.8	77536	21 AA14651
34	44.8	5.7	2268	21 AA10501
35	44.8	5.7	2618	21 AA38760
36	44.8	5.7	3252	21 AA10505
37	44.6	5.7	2376	20 AA16153
38	44.6	5.7	5324	22 AAC90079
39	44.6	5.7	11320	21 AA287298
40	44.6	5.7	12588	15 AA063293
41	44.6	5.7	36778	21 AA287318
42	44.6	5.7	37948	21 AA287285
43	44.6	5.7	38506	21 AA275633
44	44.6	5.7	38506	21 AA256001
45	44.4	5.6	867	22 AA486828

#### ALIGNMENTS

AAV72359	1	AAV72359 standard; DNA; 1116 BP.
XX	XX	AAV72359;
XX	XX	28-JUL-1999 (first entry)
DE	XX	Actinomyces sp. 36kD cellulase DNA.
XX	XX	Cellulase; detergent; animal feed; nutritional value; textile;
KW	XX	stone washing; texture modification; appearance; cellulosic fabric;
KW	XX	pulp; draining; paper; baking additive; starch treatment; grain;
KW	XX	high-fructose corn syrup production; ethanol production; fibre reduction;
XX	XX	milling; ss.
OS	XX	Actinomyces sp.
XX	XX	WO925846-A2.
XX	XX	27-MAY-1999.
PF	XX	18-NOV-1998; 98WO-US24649.
XX	XX	24-JUN-1998; 98US-0104308.
PR	XX	19-NOV-1997; 97US-0974041.
PR	XX	19-NOV-1997; 97US-0974042.
PA	XX	(GENV ) GENENCOR INT INC.
XX	XX	Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;
PI	XX	

Mycobacterium tube  
M. tuberculosis an  
M. tuberculosis re  
M. carbonacea DNA  
Mycobacterium tube  
Streptomyces feno  
Streptomyces roseo  
Total DNA sequence  
S. aureofaciens DN  
S. lividans protea  
Tripeptidyl aminop  
cDNA sequence enco  
S. venezuelae vep  
Pimaricin biosynth  
Polyanin brachys  
Platanoid synthet  
Platanoid synthet  
Myceliophthora the  
Mycobacterium tube  
M. tuberculosis im  
Mycobacterium tube  
M. tuberculosis an  
M. tuberculosis re  
Nucleotide sequenc  
Non-reducing sacch  
Rice 1-deoxy-D-xy  
Non-reducing sacch  
Mouse Sox1 cDNA.  
L05390 cDNA clone.  
S. venezuelae macr  
Sequence encoding  
S. venezuelae pik  
S. venezuelae pik  
Nucleotide sequenc  
Recombinant cosmid  
H. vitreusens inorg

DR MPI: 1999-347481/29.  
 DR P-PSDB: AAY08473.

PT New Actinomycete cellulase useful in detergent compositions, in  
 PT animal feeds and in treatment of textiles

PS Claim 4; Fig 2; 36pp: English.

CC This invention describes a novel cellulase isolated from an Actinomycete  
 CC sp. which can be used in detergent compositions, as animal feeds (to  
 CC increase nutritional value) and in treatment of textiles (e.g. stone  
 CC washing or modifying texture, feel and/or appearance of cellulosic  
 CC fabrics, including removal of 'immature' or 'dead' cotton), pulp (to  
 CC improve draining) and paper. They may also be used as baking additives,  
 CC for treating starch (in production of high-fructose corn syrup or  
 CC ethanol) and for treating grain (to reduce fibre during milling).

SO Sequence 1116 BP; 191 A; 432 C; 340 G; 153 T; 0 other;

Query Match 16.8%; Score 131.8; DB 20; Length 1116;  
 Best Local Similarity 50.7%; Pred. No. 1.5e-18;  
 Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

QY 98 ctgagcccgagcgacgagctgagctgtgagcgtgagcgcgcgagatgtgcccggg 157  
 DB 77 cggctcccgcccgagcaacagatctgcagccgtacagcaccacacagatccag 136  
 QY 158 gggcgaccgggtgatacaacagctatggggcgcgagagacccagtgatggagtcg 217  
 DB 137 accggtaagtggtgcagaaacacgctggggcgacagcgccacacagtgataatgta 196  
 QY 218 gactggaacgggcaacttcacgatacagggcgagatcacgacaacggcaacgttg 277  
 DB 197 ccggcaacggttcgagatcacccagcgagcggttcgtgctcgacacaaagcgcccca 256  
 QY 278 ccgcatatccgagcatatctactgtgctgacactggggcgctcagagcaatctggat 337  
 DB 257 agtccatccctcgtctacgagcggtctcagcaactcagcgcccgagcagcgc 316  
 QY 338 tgcgagcgagcggtgcagagagctgtccgagctgcgacagagctgagacgtacgagca 397  
 DB 317 tgcgcatgagatcagatcgatcgagcgcgcccgagctgtctctacgctacacg 376  
 QY 398 cgaagcgcgctggaatgagcgtacgacatctgtgtcagtcctccgtacgaattccgga 457  
 DB 377 gcaacggcgctctacaacgacgctacgacatctg-----ctggaacccgacaccccgca 430  
 QY 458 agcgctacagcgcgcgccgagctgatactgctgctgagtaactggaacggcggtatgc 517  
 DB 431 ccaacggggtgaaacggagacgagatcatgctgtgttcaaccgggtcggcccgctcagc 490  
 QY 518 cggcgagcgagcgcgctgcgacacgctggaactgagcggggacacctggaagttgtatg 577  
 DB 491 ccatcggttcgcgcgctgcgacgcccacagctcgcgccgacgactgggagttgagccg 550  
 QY 578 cgaactggagctggaattacatcgtcctacccggcgacgagacccacacgctcgagcg 637  
 DB 551 gcaagcaacggttcgaacgagctgattctctctggtgcctctcgcgactaagcagctgga 610  
 QY 638 agtggagctgagagcttcacagagagcggttgcgcccggagctacatccgcccggag 697  
 DB 611 gacttcagctgaagagcttcgtgacacagggcgctcagcgacggctgacacccggagct 670  
 QY 698 ggtatctgatacggtgagagacgggcttcgaaactctgggagagggcgggcgctcgcga 757  
 DB 671 ggaactcaacagatcacagcggttcgagcgctggagagggcgacacggctctggccg 730  
 QY 758 ggcgagatttttcgtaacgggtga 782  
 DB 731 tgaactcgttctcctcgcggtgaa 755

RESULT 2

AA257029  
 ID AA257029 standard; DNA; 1116 BP.

AC AA257029;

DT 19-MAY-2000 (first entry)

DE Actinomycetes cellulase protein encoding DNA.

KW Cellulase; Actinomycetes; detergent; feed additive; textile treatment;

KM pulp; paper; ds.

OS Streptomyces sp.

FH Key Location/Qualifiers

FT CDS 1.1116

FT sig\_peptide 1.81

FT mat\_peptide 82.1113

PN WO200009707-A1.

PD 24-FEB-2000.

PF 28-MAY-1999; 99WO-US11971.

PR 24-JUN-1998; 98US-0104308.

PR 18-NOV-1998; 98WO-US24649.

PR 28-MAY-1999; 99US-0321981.

PA (GENEW) GENENCOR INT INC.

PI Jones BE, Van Der Kleij WH, Van Solingen P, Weyler W;

DR MPI: 2000-224344/19.

DR P-PSDB: AAY67496.

PT A novel Actinomycetes cellulase and related DNA, useful for detergent

PT compositions, treating textiles and paper or pulp

PS Claim 5; Fig 2; 72pp: English.

CC The invention provides a cellulase from Actinomycetes. The cellulase can

CC be used in a detergent composition, as an additive for animal feed and

CC for the treatment of textiles or pulp and paper. The DNA encoding the

CC cellulase can be used to identify homologous cellulases and for

CC recombinant production of cellulases. The present sequence represents a

CC DNA encoding a cellulase from Actinomycetes.

SO Sequence 1116 BP; 191 A; 432 C; 340 G; 153 T; 0 other;

QY 98 ctgagcccgagcgacgagctgagctgtgagcgtgagcgcgcgagatgtgcccggg 157  
 DB 77 cggctcccgcccgagcaacagatctgcagccgtacgacacacagatccag 136  
 QY 158 gggcgaccgggtgatacaacagctatggggcgcgagagaccccgatgagtgagtcg 217  
 DB 137 accggtaagtggtgcagaaacacgctggggcgacagcgccacacagtgataatgta 196  
 QY 218 gactggaacgggcaacttcacgatacagggcgagatcacgacaacggcaacgttg 277  
 DB 197 ccggcaacggttcgagatcacccagcgagctgctgctcgacacaaagcgcccca 256  
 QY 278 ccgcatatccgagcatatctactgtgctcactggggcgcccgacgagcaatctggag 337

```

Db      257  agtctatccctcggtctagacagcgctgcacactacggaactgctgccccgcagcagc 316
Qy      338  ttgcgcgcgcgtgacagactgtgcagcgtgcagacgttgagactcaagcgcgtca 397
Db      317  ttgccatgcgcatcagctcatcgccagcgcgcagcagcatgtctctacgctacacg 376
Qy      398  cgaacggcgcgtggaatgcgcgtacacatctggttcagttccctacgaattccggca 457
Db      377  gcaacgcgtctacaacgcgcgtgacgaattctg-----ctggaccgcaaaccccgca 430
Qy      458  acggtctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 517
Db      431  ccaacggggtgacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 490
Qy      518  cgggcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 577
Db      491  ccatcggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 550
Qy      578  ccgactggagctggaattacatcgctcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 637
Db      551  gcagcaacgcgttcgaacgcgtgtatctctctctcgcgcgcgcgcgcgcgcgcgcgcgc 610
Qy      638  agctgacccgaagcgccttatcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 697
Db      611  gcttcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 670
Qy      698  ggtatctcgtcgtgtagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 757
Db      671  ggtaccctcacacgacatcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 730
Qy      758  gcgcgcgatttttcgcgtacgcgtgca 782
Db      731  tgaactcgtctctctccgcgcgcgtgca 755

RESULT      3
AA257031    ID      AA257031 standard; DNA: 1715 BP.
XX
AC      AA257031:
DT      19-MAY-2000 (first entry)
XX
DE      DNA sequence of an expression cassette containing cellulase 11AG8.
XX
KM      Cellulase; Actinomycetes; detergent; feed additive; textile treatment;
OS      pulp; paper; cellulase 11AG8; ds.
XX
OS      Streptomyces lividans.
XX
XX      Synthetic.
XX
FH      Key
FT      Location/Qualifiers
FT      1..406
FT      promoter
FT      /tag= a
FT      /note= "GI promoter sequence"
FT      407..548
FT      sig_peptide
FT      /tag= b
FT      /note= "Cela signal sequence"
FT      549..1564
FT      mat_peptide
FT      /tag= c
FT      /note= "mature 11AG8 sequence"
FT      1576..1715
FT      misc_feature
FT      /tag= d
FT      /note= "11AG3 downstream sequence containing terminator"
XX
PN      MO200009707-A1.
XX
PD      24-FEB-2000.
XX
XX      28-MAY-1999; 99MO-US11971.
PF
XX      24-JUN-1998; 98US-0104308.
PR

```

```

PR      18-NOV-1998; 98MO-US24649.
PR      28-MAY-1999; 99US-0321981.
XX
PA      (GEMV ) GENENCOR INT INC.
XX
PI      Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;
XX
DR      WPI: 2000-224344/19.
DR      P-PSDB; AAY67497.
XX
PT      A novel Actinomycetes cellulase and related DNA, useful for detergent
XX      compositions, treating textiles and paper or pulp
XX
PS      Example 6; Fig 15; 72pp; English.
XX
CC      The invention provides a cellulase from Actinomycetes. The cellulase can
CC      be used in a detergent composition, as an additive for animal feed and
CC      for the treatment of textiles or pulp and paper. The DNA encoding the
CC      cellulase can be used to identify homologous cellulases and for
CC      recombinant production of cellulases. The present sequence represents
CC      the DNA sequence of a complete expression cassette consisting of the GI
CC      promoter, cela signal sequence, cellulase 11AG8 and GI terminator
XX
SQ      Sequence 1715 BP; 306 A; 612 C; 558 G; 239 T; 0 other:

Query Match      16.7%; Score 131.6; DB 21; Length 1715;
Best Local Similarity 50.7%; Pred. No. 1.7e-18;
Matches 345; Conservative 0; Mismatches 329; Indels 6; Gaps 1;

Qy      103  ccgaagccgcgcgtcgacgtgtcgagcgcgtggacgcgcgcgcgcgcgcgcgcgcgcgcgc 162
Db      533  ccgcacaagccacaacccagcagatctgcaccgtctacgcaccaccagcaccgcgcgcgcgc 592
Qy      163  taccgggtgatcaacaacgatactgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 222
Db      593  taccgtgtgacagaacaacgcgcgtgggcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 652
Qy      223  gaaacgggcaactcaagatcacacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 282
Db      653  aacggttcgagatcaaccacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 712
Qy      283  taccgcgcatactactcgtgtgcacactgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 342
Db      713  tatccctcgtctacagcgcgttcaccatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 772
Qy      343  cgcgcgcgtgaggaagcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 402
Db      773  atcgagatcagctcgatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 832
Qy      403  ggcgcgtggaatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 462
Db      833  ggcgtctacaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 886
Qy      463  tacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 522
Db      887  ggggtggaacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 946
Qy      523  ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 582
Db      947  ggttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1006
Qy      583  tgggactggaattacatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 642
Db      1007  aacggttcgaacgacgagatctctctctcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1066
Qy      643  gacctgaagccttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 702
Db      1067  gacctgaagccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1126
Qy      703  ctgcatacgttgagagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 762
Db      1127  ctcaacgacatcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1186

```





FT	CDS	complement (12113..13150)	FT		
FT		/tag= 1	FT	misc_feature	/tag= aj
FT	CDS	/note= "fkbl gene"	FT		/note= "KR2"
FT		complement (13212..23988)	FT	misc_feature	complement (34606..35749)
FT		/tag= m	FT		/tag= ak
FT	misc_feature	/note= "fkbc gene"	FT	misc_feature	complement (28750..28960)
FT		complement (13452..13662)	FT		/tag= al
FT		/tag= n	FT	misc_feature	/note= "DH2 (inactive)"
FT	misc_feature	/note= "ACP6"	FT		complement (35749..37144)
FT		complement (13761..14394)	FT		/tag= am
FT		/tag= o	FT	misc_feature	/note= "AT2"
FT	misc_feature	/note= "KR6"	FT		complement (37145..38296)
FT		complement (14517..15294)	FT		/tag= an
FT		/tag= p	FT	misc_feature	/note= "KS2"
FT	misc_feature	/note= "ER6"	FT		complement (38371..38581)
FT		complement (15438..16587)	FT		/tag= ao
FT		/tag= q	FT	misc_feature	/note= "ACP1"
FT	misc_feature	/note= "dehydratase domain (DH) 6"	FT		complement (38677..39307)
FT		complement (16587..17820)	FT		/tag= ap
FT		/tag= r	FT	misc_feature	/note= "KR1"
FT	misc_feature	/note= "acylttransferase domain (AT) 6"	FT		complement (39442..40609)
FT		complement (17820..19053)	FT		/tag= ag
FT		/tag= s	FT	misc_feature	/note= "DH1"
FT	misc_feature	/note= "KS6"	FT		complement (40609..41842)
FT		complement (19116..19326)	FT		/tag= ar
FT		/tag= t	FT	misc_feature	/note= "AT1"
FT	misc_feature	/note= "ACP5"	FT		complement (41842..43093)
FT		complement (19464..20097)	FT		/tag= as
FT		/tag= u	FT	misc_feature	/note= "KS of extender module 1 (KS1)"
FT	misc_feature	/note= "KR5"	FT		complement (43144..43660)
FT		complement (20241..21420)	FT		/tag= at
FT		/tag= v	FT	misc_feature	/note= "ACP of loading domain"
FT	misc_feature	/note= "DH5"	FT		complement (43777..44629)
FT		complement (21420..22653)	FT		/tag= au
FT		/tag= w	FT	misc_feature	/note= "ER of loading domain"
FT	misc_feature	/note= "AF5"	FT		complement (44974..46573)
FT		complement (22653..23892)	FT		/tag= av
FT		/tag= x	FT	CDS	/note= "COA ligase of loading domain"
FT	CDS	/note= "KS5"	FT		46734..47788
FT		complement (23992..46573)	FT		/tag= aw
FT		/tag= y	FT	CDS	/note= "fkpo gene"
FT	misc_feature	/note= "fkbb gene"	FT		47785..52272
FT		complement (24163..24373)	FT		/tag= ax
FT		/tag= z	FT	CDS	/note= "fkdp gene"
FT	misc_feature	/note= "ACP4"	FT		52275..71465
FT		complement (24997..26146)	FT		/tag= ay
FT		/tag= aa	FT	misc_feature	/note= "fkba gene"
FT	misc_feature	/note= "DH4 (inactive)"	FT		52362..53576
FT		complement (26146..27430)	FT		/tag= az
FT		/tag= ab	FT		/note= "KS7"
FT	misc_feature	/note= "AT4"	FT	misc_feature	53577..54716
FT		complement (27430..28684)	FT		/tag= bb
FT	misc_feature	/note= "ACB3"	FT		/note= "AT7"
FT		complement (27430..28684)	FT	misc_feature	54717..55871
FT		/tag= ac	FT		/tag= ba
FT	misc_feature	/note= "KS4"	FT		/note= "DH7"
FT		complement (28750..28960)	FT	misc_feature	56019..56819
FT		/tag= ad	FT		/tag= bc
FT	misc_feature	/note= "ACP3"	FT		/note= "ER7"
FT		complement (29092..29740)	FT	misc_feature	56943..57575
FT		/tag= ae	FT		/tag= bd
FT	misc_feature	/note= "KR3"	FT		/note= "KR7"
FT		complement (29869..31018)	FT	misc_feature	57710..57920
FT		/tag= af	FT		/tag= be
FT	misc_feature	/note= "DH3 (inactive)"	FT		/note= "ACP7"
FT		complement (31018..32185)	FT	misc_feature	57990..59243
FT		/tag= ag	FT		/tag= bf
FT	misc_feature	/note= "AT3"	FT		/note= "KS8"
FT		complement (32185..33439)	FT	misc_feature	59244..60398
FT		/tag= ah	FT		/tag= bg
FT	misc_feature	/note= "KS3"	FT		/note= "AT8"
FT		complement (33505..33715)	FT	misc_feature	60399..61412
FT		/tag= ai	FT		/tag= bh
FT	misc_feature	/note= "ACP2"	FT		/note= "DH8 (inactive)"
FT		complement (33823..34480)			

Query Match	Similarity	Score	DB	Length
Best Local	48.4%	Pred. No. 0.032		
Matches 149	Conservative	0	Mismatches 159	Indels 0
			Gaps	
QY 447	gaatccggcaacggcgtacagcgcgcgccagcagctgatactgctgaactggaacg	506		
DB 27680	GAAGTCCAAACATCGGCCACACATCAGGCCCGCGGTGTGCGCGGTCTATCAAGATG	27621		
QY 507	cggcgatgacgcggcgcgagcgcgctggccacccgctggaactggccggggcaactggga	566		
DB 27620	CGTCGCGCAATCGCGGACAGCAGCCCTGCGCGGACACCCGACGCGAGACAGCCGCTCGCA	27561		
QY 567	agctcgtatgcgcgactggactggaattaaatcgcctcaccggcggaagcaccacac	626		
DB 27560	CGTCGACTGAGAGCGCGCGCGCTGCAACTCTCACCAGTATGCCCGCTGCGCCCAAGAC	27501		
QY 627	gtcgtgtgacgacgtgagactgaagaccttcaatcgacgacggtgcgcgcgactaac	686		
DB 27500	CGACGCGCGCGCGCGCGCGGTGTGTCCTCTTCGCGCGTACAGCGCACACGCCACG	27441		
QY 687	ccggcgcgagatggtatctgcatactggtgagacgggcttcgaactctggagggcgggc	746		
DB 27440	CCTGCTGGAGAGCCACCCCGCGGGAGCGCGCGCGAGAGCCGCTCGCAAGCC	27381		
QY 747	cggtctgc 754			
DB 27380	CGGTGAGC 27373			
RESULT 6				
AAT16766				
ID AAT16766	standard; DNA; 756 BP.			
XX AAT16766;				
DT 24-FEB-1997	(first entry)			
XX Collagen-like polymer sequence D gene 2 encoding unit CZA24C2.				
KM collagen; repetitive triad motif; recombinant production;				
KW photographic; medical; structural; fibre; ds.				
OS Synthetic.				
XX US5496712-A.				
XX PD 05-MAR-1996.				
XX PF 06-NOV-1990;	90US-0609716.			
XX PR 05-NOV-1992;	92US-0972032.			
XX PR 06-NOV-1990;	90US-0609716.			
XX PR 12-NOV-1991;	91US-0791960.			
XX PA (PROT-) PROTEIN POLYMER.				
XX PI Cappello J, Ferrari FA;				
XX DR WPI; 1996-150728/15.				
XX DR P-PSDB; AAR93248.				
PT Collagen-like polymers comprising repetitive triads - produced in				
PT unicellular organisms with improved characteristics, useful in, e.g.				
PT photographic and medical fibres				
XX Example 2; Column 17-18; 43pp; English.				
XX The invention concerns collagen-like polymers having repetitive triads				
CC with reduced proline content, and where glycine is the initial amino				
CC acid and the subsequent amino acids are varied. The choice of triads				
CC utilised in a recombinant collagen-like polymer are chosen in order to				
CC affect properties such as helix stability, hydration, solubility, gel				

Query Match	Best Local Similarity	Score	DB 17:	Length	756:
Matches	222: Conservative	0:	Mismatches 287:	Indels	0:
123	gtgcgaagcctgaggacgcgcgcgcatgtgcccggggggcgccctacccgggtgatcaacaagt	182			
218	gtgcgccttgaccgctggttccacggggtgcctccgggacctgagcccgccagtgctgcgc	277			
183	atgvggcggcgaggaacccgacagltgcatltgagtlcggaactggaacgggcaacttcagat	242			
278	ctggagacactggttccacccgggtgtctccgggacctgagcccgccaggtgctgccttgagc	337			
243	caacaaggccgcatcaacgaacaaacgcaacaactgtgcccgcctatccggcatctacttcgg	302			
338	cggctgtgtccaccgggtgtctccgggacctgacaggcccgccaggtgtcgcttgagccggcttg	397			
303	gtgcacactgtgggcgccttgacagcaacatctcggaattgtccgcggcgctgtcaggaagctgtc	362			
398	gtccacccgggtgtctccgggacctgacagccgcgcgcaggtgtgccttgagaccgctgttcaac	457			
363	cgaactgtgcagagactgtgaagctcaacgcgcgatcaacgaacggggcgcgtctgtgaatgcgcta	422			
458	cgggtgtctccgggacctgagggcccgccagggtgtgcctctggagccggtgttccacccggctg	517			
423	cgacatctgttccagtcaccggtcacgaattctcggcaacggcctacagcgcgcgccgagct	482			
518	ctccgggacctgacagcccgccaggtgcgcctctggaacggctgtgttccacccgggtgtccgg	577			
483	gatgatctggtcgtgaactgtgaacggcgcggtgatatgccgggcggcagccgctgtgcccagct	542			
578	ggaactgtcgaagcccgccaggtgtgcgcctggacccggctgtgtccacccgggtgtccgggacctg	637			
543	ggaactgtcggcgccacactgtggaaactctgtgatatgtaacgcgactggaactgaattacatgc	602			
638	caggcccgccaggtgtgcgcttggaacggcctgtgttccacccgggtgtctccgggacctgtcagacc	697			
603	ctacccggcgacgaagcccccacacgctcgg	631			
698	cgccaggtgtctacagccccagcaggtcgg	726			

RESULT 7  
 AAA38763  
 ID AAA38763 standard; cDNA, 1404 BP.  
 AC AAA38763;  
 XX  
 XX  
 DT 14-SEP-2000 (first entry)  
 XX  
 XX  
 DE Wheat 1-deoxy-D-xylose 5-phosphate synthase putative coding sequence.  
 XX  
 KW Wheat; 1-deoxy-D-xylose 5-phosphate synthase; DXPS;  
 XX isoprenoid biosynthesis; herbicide; ss.  
 OS Triticum aestivum.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 31..1404

```

FT      /tag- a
FT      /product- "1-deoxy-D-xylose 5-phosphate synthase"
FT      /partial
XX      WO200032792-A2.
XX      08-JUN-2000.
XX      02-DEC-1999; 99WO-US28587.
XX      03-DEC-1998; 98US-0110779.
XX      (DUPO ) DU PONT DE NEMOURS & CO E I.
XX      Cahoon RE, Tao Y, Williams ME, Coughlan SJ, Weng Z;
XX      P-PsDB; AAY97425.
XX      MPI: 2000-412338/35.
XX      P-PsDB; AAY97425.
XX      Polynucleotide encoding 1-deoxy-D-xylose 5-phosphate synthase enzyme
XX      useful for producing transgenic plants and for producing antibodies
XX      specific to which is useful for screening cDNA expression libraries
XX      Claim 2: Page 67; 73pp; English.
XX      The present sequence is a putative coding sequence for the wheat
XX      1-deoxy-D-xylose 5-phosphate synthase enzyme (DXPS). Its protein is
XX      involved in the isoprenoid biosynthesis pathway. The cDNA was identified
XX      by sequencing a number of clones and then comparing their protein
XX      sequences to known proteins; this showed the sequence's similarity to the
XX      Capsicum annum DXPS sequence. The DXPS gene and protein can be used to
XX      create transgenic plants which express the gene at either different
XX      levels or at different stages of development compared to normal, and to
XX      identify herbicides.
XX      Sequence 1404 BP; 277 A; 482 C; 439 G; 206 T; 0 other;
SQ
Query Match 6.3%; Score 49.4; DB 21; Length 1404;
Best Local Similarity 46.9%; Pred. No. 0.18;
Matches 220; Conservative 0; Mismatches 246; Indels 3; Gaps 2;
XX      229 ggaacttcagatcacacggcgatcagacaagcagcaaacgagcgactacccg 288
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      355 ggcgtcgtcagatcacacggcgatcagacaagcagcaaacgagcgactacccg 414
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      289 ggcactactcgggtggtcagctggcgccctgcagcagcaatcgagatggcgcgcg 348
XX      | || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      415 ctctggagcgtgcgcaaccgctgtaccgcagacaagattctgacggggcgcgcgataag 474
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      349 gtgcaggaagctgtcgcagctgcagcagagcttgaacgtcacgcagatcacgagcgcg 408
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      475 atgcgacagatg--cggcagaccacagcgtgtcgcgttcctcgaagcgctcagag-agg 531
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      409 tggaaatgcgcgtcagacatctgttcagctccggttcacagatctccgcgcaacgctaacg 468
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      532 gagtacgacagcgtcgcgacccgacacagctcacacacatctccgcgcgcctcgagatg 591
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      469 gggcgcgccagctgagatctgctgaacttgaaacggcgcgctgagtcggcgcgagc 528
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      592 ggcgtcggggaggaacctcaaggcgcgaaacaacgtgtgtgcggtgagatgtgggagcg 651
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      529 cgcgttcgacacgttggaactgtccggggcgcaaccttggaagctgtgatatgcgaactggag 588
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      652 gccatgacggcgcgagcggtacgagcgatgaaacacgcggtaccctcgcagctcgagc 711
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      589 tggaaattacatgcgtaccacggcgagcagcaccacacgctcgtgtgagcgagctggagactg 648
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      712 atgactcgtgactctcaacgacaacgagcaggtgtcgtctgcgacggcgagcagcgtcgagcg 771
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      649 aagcgcttcacagcagcgtgtccgcggcgatcacatccgcgcgaggt 697
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      772 ccggcgccgcgcgtgtggcgcgctcagcagcgccctcagcagatgtcaggt 820

```

```

RESULT 8
AAQ43032
ID      AAQ43032 standard; DNA; 756 BP.
XX      AC      AAQ43032;
XX      DT      07-SEP-1993 (first entry)
XX      DE      Collagen-like polymer DCP3-C2(AB12)C2 coding sequence.
XX      KW      Recombinant; collagen-like polymer; CLP; tripeptide; helix; membrane;
XX      KW      fibre; film; coating; triad sequence; collagen; mammalian; moulding;
XX      KW      hydrogel; interchain linkage; colloid suspension; DCP; antibody; ds.
XX      OS      Synthetic.
XX      PN      WO9310154-A.
XX      PD      27-MAY-1993.
XX      PF      04-NOV-1992; 92WO-US09485.
XX      PR      12-NOV-1991; 91US-0791960.
XX      PA      (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX      PI      Cappello J, Ferrari FA;
XX      DR      MPI: 1993-182496/22.
XX      DR      P-PsDB; AAR37738.
XX      PT      High mol. wt. collagen-like protein polymers - capable of being
XX      PT      produced in unicellular microorganisms
XX      PS      Disclosure; Page 39; 82pp; English.
XX      CC      The sequences given in AAQ43030-34 encode examples of recombinantly
XX      CC      produced DCP collagen-like polymers (CLPs) which consist of repeated
XX      CC      tripeptide sequences selected from a wide range of GXY sequences,
XX      CC      where X and Y can be any amino acid. These sequences can be cloned
XX      CC      into plasmids and used to transform E. coli to produce the DCP
XX      CC      proteins. DCP peptides comprise repeated units of: A = GAGGPGGP,
XX      CC      B = GSKDPPGP and/or C = GAGGPGPK. These polymers may be used to
XX      CC      raise anti-DCP antibodies in rabbits. These polymers have molecular
XX      CC      weights of >30 kD and are able to form helices due to interchain
XX      CC      linkages. These polymers pref. contain a proportion of tripeptide
XX      CC      triad sequences found in natural collagens, pref. mammalian collagens.
XX      CC      The CLPs impart unique characteristics to materials such as fibres,
XX      CC      membranes, films, coatings, hydrogels, colloid suspensions and moulded
XX      CC      articles.
XX      SQ      Sequence 756 BP; 84 A; 288 C; 291 G; 93 T; 0 other;
Query Match 6.2%; Score 49; DB 14; Length 756;
Best Local Similarity 43.5%; Pred. No. 0.21;
Matches 223; Conservative 0; Mismatches 290; Indels 0; Gaps 0;
XX      123 gtccggaacgttggaagcgcgcgcatgtgcccggggcgctaccaggtgtacacaacgt 182
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      218 gtgcgcttggaacggcgtgtgtccacgggtgtctccggagcttgagagcccgccaggtgcgc 277
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      183 atggggcgcggaagcccgccagtgcatgtgagctgcggaactggaacgggcaactcaacgat 242
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      278 ctggaccggtgtgtccacggcggtgtcccggaaccttgagcccgcaaggtgtgcctggagc 337
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      243 caacgggacgcatcacagcagcaacaacgtgtgcgcgctacatccgcatctacttcg 302
XX      | || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      338 cgcgtgtgtccacggggtgtcgcggagcttcgagcccgcaaggtgtgcctgcgcggctcg 397
XX      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX      303 gtccacttgggcgccctgtcacgagcaattccggatgtccggcgcgctgtcaggaagctgtgc 362

```



PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 and diagnosis of tuberculosis

XX Claim 4; Page 187-188; 250pp; English.

XX This is the 3' region of DNA coding for an antigenic portion of  
 CC Mycobacterium tuberculosis antigen XP25; 5' DNA is provided in  
 CC AAV4438. XP25 DNA was isolated from a M. tuberculosis strain Erdman  
 CC genomic DNA expression library using sera from patients having  
 CC extrapulmonary tuberculosis. It bears no similarity to known  
 CC sequences. The invention relates to methods and compositions for  
 CC diagnosing tuberculosis. It provides polypeptides (see  
 CC AAW64291-W64379) comprising an antigenic portion of a soluble M.  
 CC tuberculosis antigen, or an immunogenic portion of a M.  
 CC polypeptides, recombinant expression vectors and transformed or  
 CC transfected host cells. Also claimed are methods and diagnostic  
 CC kits for detecting M. tuberculosis infection in a patient using  
 CC these polypeptides, antibodies or oligonucleotide probes and  
 CC primers.

XX Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 6.2%; Score 48.8; DB 19; Length 985;  
 Best Local Similarity 44.9%; Pred. No. 0.24;  
 Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

OY 64 ttcccgatgagcagcaagaaagaaacgagctgagccgagccgagccgagctg 123  
 DB 175 ttcccgatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 234  
 OY 124 tgcgagcagctgagcagcagcagcagcagcagcagcagcagcagcagcagc 183  
 DB 235 accggtacccgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 294  
 OY 184 tgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 243  
 DB 295 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 354  
 OY 244 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 303  
 DB 355 ggccttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 414  
 OY 304 tgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 359  
 DB 415 atcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474  
 OY 360 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 419  
 DB 475 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 533  
 OY 420 ctacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 479  
 DB 534 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 593  
 OY 480 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 539  
 DB 594 gggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653  
 OY 540 cgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 599  
 DB 654 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 713  
 OY 600 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653  
 DB 714 cactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 767

RESULT 11  
 AA219349  
 ID AA219349 standard; DNA; 985 BP.

XX AA219349;  
 AC 05-NOV-1999 (first entry)  
 DT  
 XX M. tuberculosis antigen 3' XP25 DNA sequence.

DE Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test; ss.

XX Mycobacterium tuberculosis.

XX WO942076-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03268.

XX 05-MAY-1998; 98US-0072967.

XX 18-FEB-1998; 98US-0025197.

XX (CORI-) CORLXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skelky YAW, Twardzik DR, Vedvick JS;  
 XX WPI; 1999-527409/44.

PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX Claim 4; Page 171-172; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to  
 CC AA219225 are used in the exemplification of the present invention.

XX Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 6.2%; Score 48.8; DB 20; Length 985;  
 Best Local Similarity 44.9%; Pred. No. 0.24;  
 Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

OY 64 ttcccgatgagcagcaagaaagaaacgagctgagccgagccgagccgagctg 123  
 DB 175 ttcccgatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 234  
 OY 124 tgcgagcagctgagcagcagcagcagcagcagcagcagcagcagcagcagc 183  
 DB 235 accggtacccgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 294  
 OY 184 tgggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 243  
 DB 295 gccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 354  
 OY 244 aacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 303  
 DB 355 ggccttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 414  
 OY 304 tgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 359  
 DB 415 atcaacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 474

QY	360	gTcngacngTcngcacngcngcTngnagcTccagccagatccagacagcngccngccngcTngagTcngc	413
Db	475	gTcngcacngTcTcngccagcaaaacngcngngTccngcngTngacngTcnggcT-cngTngcngcngc	533
QY	420	cTcacagacacTcTngTcTcngTccccngTccagaaTccTccngcngaaacngcTcacacngcngcngcngc	479
Db	534	cngcacacngcngcacacngcngcngTcTcngccTngacacagacaaagcTccngcngcngcngcngcngc	593
QY	480	gTcngTngacTcTcngcTngacTngaaacngcngcngTngacTngcngcngcngcngcngccngcngcngcngc	559
Db	554	gngcacacngcngcacacngcngcngcngcngcngcngcngcngcngcngcngcngcngcngcngcngcngc	653
QY	540	cngTngaaacTngccngcngcngcacacTngngagacTcngTngTngacTngcngcngcngcngacTngaaTcac	599
Db	654	cngcngcngcacngcngTngTngcngcngcngcacagcngcngcngcngcngcngcngcngcngcngcngcngc	713
QY	600	cngcTcacngcngcacngcacngccacacngTcngTngacngagcTngacTngaaagc	653
Db	714	gacccTngacTcngccTcacngcngccacccngngaaagccagatccacacacngcngcngcngcngcngc	767

## RESULT 12

AAZ19137  
ID AAZ19137 standard; DNA; 985 BP.

AC AA219137;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen DNA encoding 3' XP25.

KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
KW vaccine; immunity; ss.

KW vaccine; immunity; ss.

05 *Mycobacterium tuberculosis*.

PN W09942118-A2.

PD 26-AUG-1999.

PF 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-0072596.

XX

PA (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodés MJ, Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;

DR WPI; 1999-527416/44.

PT New polypeptide comprising antigenic portions of *M. tuberculosis*

PS Claim 4; Page 217; 323pp; English.

This invention describes novel recombinant antigens and their encoding nucleic acids derived from *Mycobacterium tuberculosis*. The novel polypeptides are useful for detecting *M. tuberculosis* infection in a biological sample by detecting antibodies which bind with the polypeptides, and are useful as vaccines for immunizing against *M. tuberculosis* infection. The new detection methods are needed as current vaccination strategies do not provide 100% immunity.

Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query match	6.28;	Score 48.8;	DB 20;	Length 985;
-------------	-------	-------------	--------	-------------

Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

64 ttcccgatggcgaacggaagacgcgagcctgagcccgagccgacgcgtcgagctg 123

[illegible]

### RESULT 13

ID AAS08699 standard; DNA; 12152 BP.

AC AAS08699;

DT 26-SEP-2001 (first entry)

DE M. carbonacea DNA encoding Everninomycin biosynthetic enzymes.

XX

KW fermentation; ds.

05 Micromonospora carbonacea var. africana.

Key	Location/Qualifiers
FH	

FT / \*tag a

FT	RBS	complement (1069..1073)
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

	FT	CDS	/day - D
E1			1184 : .2767

```

E1      /cay = c
ET      /product= "ORF2"

```

```
complement (2683...3/33,
/*aa= q
```

```

FI      /product= "ORH3"
FT      complement /3766 42761
CDS

```

```

ET      /*tag= e
ET      /product= "OBEA"

```

```

RBS      complement (4280...4284)
ET       /ttag= f
ET

```

FT	CDS	complement (4526..5368)
FT		4526..5368

```

FT      /product= "ORF5"
FT      complement (5392..6147)
FT      /tag= h
FT      /product= "ORF6"
FT      complement (6152..6156)
FT      /tag= i
FT      complement (6194..7282)
FT      /tag= j
FT      /product= "ORF7"
FT      complement (7280..8133)
FT      /tag= k
FT      /product= "ORF8"
FT      complement (8141..8145)
FT      /tag= l
FT      complement (8254..9318)
FT      /tag= m
FT      /product= "ORF9"
FT      complement (9324..9328)
FT      /tag= n
FT      /tag= o
FT      /product= "ORF10"
FT      9568..9571
FT      /tag= p
FT      10584..11585
FT      /tag= q
FT      /product= "ORF11"

CDS
W0200151639-A2.
19-JUL-2001.
12-JAN-2001: 2001WO-US01187.
12-JAN-2000: 2000US-0175751.
(SCHE ) SCHERING CORP.
Hosted TJ, Horan AC, Wang TX;

DR      WPI: 2001-442147/47.
DR      P-PSDB: AAU04901, AAU04902, AAU04903, AAU04904, AAU04905, AAU04906,
DR      AAU04907, AAU04908, AAU04909, AAU04910, AAU04911.
XX
XX
XX      New nucleic acid molecules encoding evernimicin pathway gene
XX      products, useful for improving yields of evernimicin, to produce new
XX      evernimicin and as probes to identify homologous sequences -
XX
XX      Example 1; Fig 12; 109pp; English.
PS
XX
XX      The sequence encodes 11 proteins comprising enzymes of the
XX      evernimicin antibiotic biosynthetic pathway. A vector comprising a
XX      M. carbonacea evernimicin biosynthetic pathway resistance gene product
XX      is useful for selecting for a transfected or transformed host cell. An
XX      integrative version of the vector is useful for introducing a
XX      evernimicin pathway gene (a bottle-neck gene) into an actinomycete of
XX      the genus Micromonospora. The DNA encoding the biosynthetic proteins is
XX      useful for synthesising novel evernimicin-related compounds, arising
XX      from modifications of the DNA sequence designed to change glycosyl and
XX      modified orsellinic acid groups contained in evernimicin, for
XX      expressing functional or mutant evernimicin biosynthetic enzyme for
XX      evaluation, diagnosis and preferably biosynthesis of evernimicin or
XX      other secondary metabolic products, improving the yield of evernimicin
XX      and to produce novel evernimicins and also as a hybridisation probe to
XX      identify homologous sequences. The encoded polypeptides are useful for
XX      combinatorial biosynthesis to generate libraries of orthomycins, e.g.
XX      evernimicin analogues/homologues and drug discovery. The
XX      DNA encoding the integrase allows for increasing a given gene dosage. The
XX      integrative vector can be used to permanently integrate copies of a
XX      heterologous gene of choice into chromosomes of different hosts and to
XX      integrate genes which increase the yield of known products or to generate
XX      novel products such as hybrid antibiotics or other novel secondary
XX      metabolites. The vector can also be used to integrate antibiotic

```

```

CC      resistance genes in order to carry out bioconversions with compounds to
CC      which the strain is normally sensitive and is thus useful in fermentation
CC      processes involving e.g. Streptomycetes antibiotics.
XX
XX      Sequence 12152 BP; 1675 A; 4502 C; 4475 G; 1500 T; 0 other;

Query Match      6.2%; Score 48.4; DB 22; Length 12152;
Best Local Similarity 45.9%; Pred. No. 0.34;
Matches 166; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

OY      338 tgcgcgcgcgcgtgtcgaagagctgtccgaagtcgcacagagcttggaagctcgaagcatca 397
DB      8021 TGTGGGGGCACTCGGCAATGAGACCCGAGACCGGAGCACTGTCTCCCGAGTGTCTCC
OY      398 cgaagggccgcgtgaatgcgcgcctacagacatctgtgtcagtcagtcaccgaattccgcga 457
DB      7961 GCGTGGGCGCTCCAGAGACGCCGAGACCTTGTGTGCGCAGCCGACACACGCGCCGCTGACG
OY      458 acggtctacagcgcgcgcgcgcgcgtgatctgtgtgaacttgaaagcgcgcgtgatgc 517
DB      7901 CCGCTTTCGACACAGCTGTCGGGGGTGGGCGCTGGAAGCCGTTGACCGGATCGCACCT
OY      518 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 577
DB      7841 TCCCGGTCCGGTTCCCGGTCCCGCAACGCCCGGACGACGACGACTGCGCTGCGACA 7782
OY      578 ccgactgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 637
DB      7781 TCGACGCCAGTTCTTCGTGGCGGACGACGACGCGCGCTGGGCCCCGAGAACTGGAGGCGG 7722
OY      638 agctgaactgaagagccttcacgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 697
DB      7721 AGCTGACCTGTCGCCGCCCACTACGCGAGAGTCTTCGCTGCAACCTTCGGTCCGCGG 7662
OY      698 gg 699
DB      7661 GG 7660

RESULT 14
AA199683/C
ID      AA199683 standard; DNA; 4403765 BP.
XX
XX      AA199683;
AC
XX
XX      15-JAN-2002 (first entry)
DT
XX
XX      Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
DE
XX
XX      Mycobacterium tuberculosis: strain H37Rv; strain CDC 1551; genome;
KW      variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX      Mycobacterium tuberculosis.
OS
XX
XX      US6294328-B1.
PN
XX
XX      25-SEP-2001.
PD
XX
XX      24-JUN-1998; 98US-0103840.
PF
XX
XX      24-JUN-1998; 98US-0103840.
PR
XX
XX      24-JUN-1998; 98US-0103840.
PA      (GENO-) INST GENOMIC RES.
PI      Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
XX      WPI: 2001-647261/74.
DR
XX
XX      Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX      determining the nucleotide sequence of the strain at positions in the
XX      genome corresponding to positions where M. tuberculosis strains CDC
XX      1551 and H37Rv differ -

```

XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
 PS  
 XX The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterium pathogen.  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (A199683) and  
 CC H37Rv (A199682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.  
 XX  
 SO Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;  
 Query Match 6.1%; Score 48; DB 22; Length 4403765;  
 Best Local Similarity 45.6%; Pred. No. 0.62;  
 Matches 208; Conservative 0; Mismatches 245; Indels 3; Gaps 1;  
 QY 151 gccggggggcgctaccgggtgatcaacaagctgtggcgcgagagcccgccagtcacat 210  
 DB 3963782 GTCGGCAACGGGAACTCGGCAACCAACCTGGCGAGCGGCAATACCGCAACTCCAAAC 3963723  
 QY 211 gaagtcgagctggaacagcgcaactcacatcacacggcgcgatcacagcaagcaac 270  
 DB 3963722 GTCGGGTTCCGGCAATACCGCAACCAACCTGGCGAGCGGCAATACCGCAACTCCAAAC 3963663  
 QY 271 aacgtgcccgcctatccggccatcttcgtggtgcacatggcgcgctgcagcaagcaat 330  
 DB 3963662 ATCGCGCGCGGGGAACACCGGAGAGCAACTGGGCTTCGCGCAACACGGCAATTGGCAAC 3963603  
 QY 331 tcgggttcgcggcgcgctgtgcag---agctgtccgaagttgcgcagagctgtgacgtc 387  
 DB 3963602 ATCGGTTTCGGCAACACCGGCAACCAACCTGGGCTTCGCGCAACCAACCAACGAG 3963543  
 QY 388 acgcgcatacgaacagcgccgtggaatgcgcgcctacagacatcgtgtcagttccgcaag 447  
 DB 3963542 GTGGGCAATCGGTGGGCTGAACCTCGGGCAAGCGGAACATCGCTTGTCAACTCGGGCAAC 3963483  
 QY 448 aattccggcaacggctacagcgcgcgccgagctgatactgtgctgaactggaaagc 507  
 DB 3963482 AACCAACGTGCGTTTCTTCAACTCCGGCAACCTGGGATATGCGCAACTCCAGCGAC 3963423  
 QY 508 ggcgtgatgcggggcgcgccgctggtgcacacgtggaactgycggcgccacactggaa 567  
 DB 3963422 GCGAAATGTGGCAATCGGAACCTCGGGGGCGCACCGTGGGCCCGTGTGCGGGCCATTAAC 3963363  
 QY 568 gtcgtatgcgcagctggagactgaaattacatgcgc 603  
 DB 3963362 ACCGGTTTCGGGAACCTCGGCTTCGCTCAACACGGGC 3963327  
 RESULT 15  
 AAT93095  
 ID AAT93095 standard; cDNA: 24379 BP.  
 AC AAT93095;  
 XX  
 DT 11-MAY-1998 (first entry)  
 XX  
 DE Streptomyces frenolicin gene cluster.  
 XX  
 KW Frenolicin; antibiotic; feed additive; anticoccidial;  
 KW coccidiostats; efflux pump; butyrate starter synthase;  
 KW polyketide synthase; PKS; hemiketalase; ketoreductase; cyclase;  
 KW dehydrase; ketoreductase; hydroxylase; Streptomyces roseofulvus;  
 KW ds.  
 XX

OS Streptomyces sp.  
 XX  
 XX Key  
 FT CDS  
 FT Location/Qualifiers  
 /tag= a  
 /product= 80 kDa non-membrane protein  
 /note= "gene A (specifically claimed)"  
 2945..3916  
 /tag= b  
 /product= membrane protein  
 /note= "gene B (specifically claimed)"  
 4020..4844  
 /tag= c  
 /product= protein with 6 membrane-spanning domains  
 /note= "gene C (specifically claimed)"  
 4841..6415  
 /tag= d  
 /product= ATP-binding component of ABC transporter  
 /note= "gene D (specifically claimed)"  
 6533..7183  
 /tag= e  
 /product= unknown non-membrane protein  
 /note= "gene E (specifically claimed)"  
 7344..8897  
 /tag= f  
 /product= putative efflux pump  
 /note= "gene F (specifically claimed)"  
 9164..10012  
 /tag= g  
 /product= transcription activator  
 /note= "gene G (specifically claimed)"  
 complement (10105..10621)  
 /tag= h  
 /product= translationally coupled to gene I  
 /note= "gene H (specifically claimed)"  
 complement (10618..11628)  
 /tag= i  
 /product= homologue of fabH  
 /note= "gene I (specifically claimed)"  
 11809..12066  
 /tag= j  
 /product= acyl carrier protein  
 /note= "gene J (specifically claimed)"  
 complement (12154..13209)  
 /tag= k  
 /product= acyltransferase  
 /note= "gene K (specifically claimed)"  
 13409..14686  
 /tag= l  
 /product= PKS ketocyl synthase subunit  
 /note= "gene L (specifically claimed)"  
 14767..16047  
 /tag= m  
 /product= PKS chain length factor  
 /note= "gene M (specifically claimed)"  
 16120..16371  
 /tag= n  
 /product= acyl carrier protein used by the PKS  
 /note= "gene N (specifically claimed)"  
 complement (16453..16935)  
 /tag= o  
 /product= putative hemiketal dehydrase  
 /note= "gene O (specifically claimed)"  
 17088..17903  
 /tag= p  
 /product= ketoreductase related to actIII  
 /note= "gene P (specifically claimed)"  
 17903..18898  
 /tag= q  
 /product= cyclase/dehydrase related to act VII  
 /note= "gene Q (specifically claimed)"  
 18895..19839  
 /tag= r  
 FT CDS



```

FT FT /product= cyclase/dehydrase related to activ
FT FT /note= "gene R (specifically claimed)"
FT CDS complement (1990...20907)
FT FT /tag= s
FT FT /product= oxidoreductase
FT FT /note= "gene S (specifically claimed)"
FT CDS complement (20904...22094)
FT FT /tag= t
FT FT /product= quinone-forming hydroxylase
FT FT /note= "gene T (specifically claimed)"
FT CDS complement (22505...22179)
FT FT /*tag= u
FT FT /product= not specified
FT FT /note= "gene U"

EP806480-A2.
12-NOV-1997.
02-MAY-1997; 97EP-0107329.
07-MAY-1996; 96US-0016753.
(HOFF ) HOFFMANN LA ROCHE & CO AG F.
Reeves CD, Soliday CL;
PI
XX MPI: 1997-538619/50.
DR P-PSDB: AAM34199-W34219.
XX
PT Streptomyces frenolicin gene cluster - useful for producing
PT recombinant frenolicin antibiotics
PS Claim 1: Page 40-60; 66pp; English.
XX
CC This DNA sequence comprises the Streptomyces frenolicin gene
CC cluster containing specifically claimed coding sequences (genes
CC A-U) that respectively encode 21 proteins (see AAM34199-219) involved
CC in frenolicin synthesis. The genes can be divided into 5
CC subclusters: (1) genes A, B, C, D and F encode an efflux pump; (2)
CC genes H, I, J and K encode butyrate starter synthases; (3) genes
CC L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and
CC R encode a hemiketalase, a ketoreductase and cyclases/dehydrases;
CC and (5) genes S and T encode a keto/enoyl reductase and a
CC hydroxylase. Also claimed are vectors, host cells (especially a
CC Streptomyces sp., particularly Streptomyces roseofulvus), and the
CC encoded proteins. Cells transformed using the above sequence can
CC be cultured to produce frenolicins or frenolicin precursors. The
CC precursors can be converted to frenolicins by chemical or other
CC methods. The frenolicins can be oxidised to frenolicin B, an
CC antibiotic used as an anticoccidial agent. The frenolicins can be
CC used as animal feed additives.
XX
SQ Sequence 24379 BP: 3077 A; 9792 C; 8499 G; 3011 T; 0 other:

```

```

Query Match 6.1%; Score 47.8; DB 18; Length 24379;
Best Local Similarity 47.2%; Pred. No. 0.48;
Matches 145; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

```

```

QY 387 cagcgcatcaagcagcgccgtggaatgcgcctcagaacatcgttcagtcgcgtcac 446
   || || || || || || || || || || || || || || || || || || || ||
DB 17387 caagcgggcggtccggtcgagtgacgcgcgaactcagagatccctgtggaaga 17446
   || || || || || || || || || || || || || || || || || || || ||
QY 447 gaattccggcagcgtcagcgcgccgagcagatcagatcgtctgaactggaacgg 506
   || || || || || || || || || || || || || || || || || || || ||
DB 17447 cgtcatcgccacaactcaagaagcgtctcgtgcaccgcggaggtgtcaccacggg 17506
   || || || || || || || || || || || || || || || || || || || ||
QY 507 cggcgtgatacgcgcgccgagccgctggccacccgtggaactgcccgggccaacctggga 566
   || || || || || || || || || || || || || || || || || || || ||
DB 17507 cgggatacgacggcgcgccgctcgtcgaacatcgctccacggcgccaagcaagg 17566
   || || || || || || || || || || || || || || || || || || || ||
QY 567 agtctgtatgcgactggaactgaattacatcgctaccgagcgcaagccaccac 626

```

```

DB 17567 cgtcgtcttcggcgcccccctactcgcgtcaccagaagcgctcgtcgttcaccaaagc 17626
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 627 gtcgtgagcgagctggacctgaagccttcacgaacgagcgctcgcccggtacat 686
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 17627 cctcgccctcgaactggcccggaagcgacatacgcgtcaagcgcgtctgccccgctaagt 17686
   || || || || || || || || || || || || || || || || || || || ||
QY 687 ccgagccg 693
   | | | | |
DB 17687 cgagacg 17693

```

Search completed: August 19, 2002, 13:53:47  
Job time: 14326 sec







```

? PRIOR APPLICATION NUMBER: 08/974,042
? PRIOR FILING DATE: 1997-11-19
? NUMBER OF SEQ ID NOS: 5
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 4
? LENGTH: 1716
? TYPE: DNA
? ORGANISM: Unknown
? FEATURE:
? OTHER INFORMATION: Nearest "neighbor" = Streptomyces
? OTHER INFORMATION: thermovaceus
? OS-09-321-981-4

```

```

1      MEDIUM TYPE: Floppy disk
2      COMPUTER: IBM PC compatible
3      OPERATING SYSTEM: PC-DOS/MS-DOS
4      SOFTWARE: PatentIn Release #1.0, Version #1.30
5      CURRENT APPLICATION DATA:
6      APPLICATION NUMBER: US/08/642,255
7      FILING DATE:
8      CLASSIFICATION: 435
9      ATTORNEY/AGENT INFORMATION:
10     NAME: ROWLAND, Bertram I.
11     REGISTRATION NUMBER: 20,015
12     REFERENCE/DOCKET NUMBER: A5556-3/BIR
13     TELECOMMUNICATION INFORMATION:
14     TELEPHONE: (415) 494-8700
15     TELEFAX: (415) 494-8771
16     TELEX: 910 277299 FRT UR
17     INFORMATION FOR SEQ ID NO: 50:
18     SEQUENCE CHARACTERISTICS:
19     LENGTH: 756 base pairs
20     TYPE: nucleic acid
21     STRANDEDNESS: double
22     TOPOLOGY: linear
23     MOLECULE TYPE: other nucleic acid
24     DESCRIPTION: /desc = "synthetic"
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
```

```

? APPLICANT: Reed, Steven G.
? APPLICANT: Skeiky, Yasir A.W.
? APPLICANT: Dillon, Davin C.
? TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
? NUMBER OF SEQUENCES: 241
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: SEED AND BERRY LLP
? STREET: 6300 Columbia Center, 701 Fifth Avenue
? CITY: Seattle
? STATE: Washington
? COUNTRY: USA
? ZIP: 98104-7092
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentln Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/056,556
? FILING DATE: 07-APR-1998
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Mark David J.
? REGISTRATION NUMBER: 31,392
? REFERENCE/DOCKET NUMBER: 210121.457
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (206) 622-4900
? TELEFAX: (206) 682-6031
? INFORMATION FOR SEQ ID NO: 182:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 985 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
US-09-056-556-182

```

```

Query Match      6.2%; Score 48.8; DB 4; Length 985;
Best Local Similarity 44.9%; Pred. No. 0.015;
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

```

```

QY 64 ttcccgatgagcagcaaggaaggaacgagcctgagcccgagccgagctcgagctg 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 175 ttgcgcgggtggtccgcccgcacacagtggtggtggtggtggtggtggtggtggt 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 tggcgagcgtcgagcagcgagctgctgagcggtggtggtggtggtggtggtggt 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 235 accggatcccgccgacacggcggtgtgtggtggtggtggtggtggtggtggtggt 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 tgggagcgaggaagcagccagtgatgagtgagtgagtgagtgagtgagtgagtc 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 295 gccggcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 aacggcgagcagcagcaagcagcagcagcagcagcagcagcagcagcagcagc 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 ggcgtttgacggcgacacggcggtggtggtggtggtggtggtggtggtggtggt 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 304 tggcactgggagcgtcgagcagc---aatcgagtgagtgagtgagtgagtgagtc 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 415 atmacggggcgccggcgccggcgacacggcgacacggcgacacggcgacacgggt 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 360 gtccgagctgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 475 gccgcaggtctggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 420 ctacgacatctgttcaagtcagcgaattccggaacgagtgagtgagtgagtgag 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 534 cggcaacggcgacacggcggtggtggtggtggtggtggtggtggtggtggtggt 593
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 480 gtcgagatctgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 gggcaatggcgacacggcggtggtggtggtggtggtggtggtggtggtggtggt 653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

TREATM

```

QY 540 cgtggaactgcccgggcccacctgggaagtcgtatgtccgactgggactggaatacat 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 cggcgacaggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 600 cgcctacggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 cacttgatcggttgcacccgacacgggaagcagatccacacgacgacgagtcgc 767
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-09-103-840A-2/c
? Sequence 2, Application US/09103840A
? Patent No. 6294328
? GENERAL INFORMATION:
? APPLICANT: FLEISCHMAN, Robert D.
? APPLICANT: WHITE, Owen R.
? APPLICANT: ERASER, Claire M.
? APPLICANT: VENTER, John C.
? TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
? FILE REFERENCE: 24366-20007.00
? CURRENT APPLICATION NUMBER: US/09/103,840A
? CURRENT FILING DATE: 1998-06-24
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: Patentln Ver. 2.1
? SEQ ID NO 2
? LENGTH: 4403765
? TYPE: DNA
? ORGANISM: Mycobacterium tuberculosis
? FEATURE:
? OTHER INFORMATION: CDC 1551
? OTHER INFORMATION: "n" bases at various positions throughout the sequence
? OTHER INFORMATION: represent a, t, c or g
?
US-09-103-840A-2

```

```

Query Match      6.1%; Score 48; DB 4; Length 4403765;
Best Local Similarity 45.6%; Pred. No. 0.069;
Matches 208; Conservative 0; Mismatches 245; Indels 3; Gaps 1;

```

```

QY 151 gccggggggcgctacccggtgtgatacaagaatgtggtggcgagagccgacgtgcatc 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3963782 gttcgacacagggagacctcgacacacacacacacacacacacacacacacac 3963723
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 211 gaagtcgactggaacgggcaacttcacgatacaacggcgagcagcagcagcagcagc 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3963722 gttcggttgggaatracccgacacacacacacacacacacacacacacacac 3963663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 271 aacgtgagcgcctatccgacatctacttcggtggtggtggtggtggtggtggtggt 330
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3963662 atcgcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 3963603
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 331 tcggatctgcggcggtggtggtggtggtggtggtggtggtggtggtggtggtggt 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3963602 atcggttgcgacacacacacacacacacacacacacacacacacacacacac 3963543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 388 acgcgacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 447
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3963542 gttggcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 3963483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 448 aattcggaacggtctacagcagcagcagcagcagcagcagcagcagcagcagcagc 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3963482 aacacagtcggttcttcaatccgacacacacacacacacacacacacacacac 3963423
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 508 ggcgtatgcggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3963422 cggaaatgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 3963363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 568 gtctgtatgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 603
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3963362 accggttttcgggaactcggttgcgtcaacacggggc 3963327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```











```

Db 459 GCGCTGTSGCGSRGRCGCGSGS---CTSGACGYSWSTGCAACACCGCGGGGCTTG 515
Qy 280 gcatatccgcatctactctcgtggtccacttgaggcgctgacagcaatccgattg 339
Db 516 GCGCGGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 575
Qy 340 ccgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 399
Db 576 GTCCTCCGCGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 635
Qy 400 agggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 459
Db 636 ACCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 694
Qy 460 ggcatacagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 519
Db 695 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 754
Qy 520 ggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 579
Db 755 AACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 814
Qy 580 gactgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 639
Db 815 GTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 874
Qy 640 ctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 699
Db 875 GACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 934
Qy 700 tctcgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 746
Db 935 AAGATGAGAGAGTGTCTCAGGGGCGCTGCGCGCGCGCGCGCGCGCGCGCG 981

RESULT 14
US-08-804-227C-7/c
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4437 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7

```

```

Query Match 5.9%; Score 46; DB 2; Length 44377;
Best Local Similarity 47.6%; Pred. No. 0.1;
Matches 136; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```

```

Qy 70 gatggcgagcaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 129
Db 18968 GACGGGGGGGTGGCGCTTGGCGCGGAGCACCGTGTGATCCCTCGCGCGTGGGCGGA 18909
Qy 130 cgtggaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 189
Db 18908 GACGACCGCGTGTGTGAGGACCGCGCGGTGTGACACGACGAGTCAGCGGCTTTCGTC 18849
Qy 190 gcggaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 249
Db 18848 GGGGAGTGGCTGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18789
Qy 250 gccgatcaagcaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 309
Db 18788 GCGGACCTCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18729
Qy 310 tggggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 355
Db 18728 GGGACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18683

```

```

RESULT 15
US-08-804-198-1/c
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:

```

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-198-1

Query Match 5.9%; Score 46; DB 2; Length 4437;  
Best Local Similarity 47.6%; Pred. No. 0.1;  
Matches 136; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
QY 70 gatggcgacaacggaagaaagacccgagccttgagcccgagccgacgctcgagctgtgcgga 129  
DB 18968 GACGGGCGGCTCGCTCGCTGGCGGAGCACCGGTGTCATCCCTCGCGCTTGGGCGGA 18909  
QY 130 cgtctggagacgagcgagatgtgagcgaggggagcgtacacgggtgatacaacaagatagggc 189  
DB 18908 GAGCAGCGCGCTCGTCGAGCACCCCGGCGGTGTGACACGACGACGACGCGGTGTCTC 18849  
QY 190 gcggagacgagccagtgcaatlgaggtcgagacttggaaacgagcaacttcaagatcacagcg 249  
DB 18848 GGGGAGTGGGTCTGAGCAGGCGGCGGCGGCTCGGTGACGTCACAGGCGCCGTAC 18789  
QY 250 gccgatacacgagcaacgagcaacaaagcgtgcgacctatccggccatctcattcggtgccaac 309  
DB 18788 GCGGACCTCGGACACCGACTCGGTCAACTGGCGACCAAGTCGATGACGCGCGGTGCGTC 18729  
QY 310 tggggcgagctgcagagcaatttcgggattgcgcgagcggtgcag 355  
DB 18728 GGGACCGCGCTGCCACCAAGAGATGGCGCACCTGTGTCCGG 18683

Search completed: August 19, 2002, 13:45:36  
Job time: 13940 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 10:59:49 ; Search time 3821.57 Seconds  
(without alignments)  
2775.981 Million cell updates/sec

Title: US-10-003-759-3  
Perfect score: 786  
Sequence: 1 atgaacgtcatgcgtgcgtl.....ttccgtacggtgcagtag 786

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estinu:\*  
5: em\_estlov:\*  
6: em\_estlpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No	Score	Query Match	Length	ID	Description
C 1	60.8	7.7	925	12	CNS0091P
2	56.2	7.2	925	12	CNS0091P
3	55.6	7.1	932	12	CNS00720
4	54.4	6.9	864	10	BF253875
5	53.8	6.8	964	12	CNS003WG
6	50.6	6.4	700	12	AG060149
7	50.4	6.4	935	12	CNS006XK
8	50.4	6.4	1022	9	BE040131
9	50	6.4	894	12	AG060214
C 10	50	6.4	1009	12	CNS010EW
C 11	49.8	6.3	449	10	BE517742
C 12	49.6	6.3	682	9	AU095988
C 13	49.2	6.3	534	12	AZ934730
C 14	49.2	6.3	546	12	AZ935386
C 15	49.2	6.3	644	9	AU030637
16	48.2	6.1	443	9	AV914669
17	48	6.1	446	10	BM373545

18	48	6.1	559	10	BE405069	BE405069	WHE1215-G
19	48	6.1	588	10	BF255157	BF255157	HVSMF000
20	48	6.1	604	9	AV941745	AV941745	AV941745
21	48	6.1	606	10	BF255408	BF255408	HVSMF000
22	48	6.1	631	9	AL505254	AL505254	AL505254
23	48	6.1	635	9	AV932837	AV932837	AV932837
24	48	6.1	637	9	AV933847	AV933847	AV933847
C 25	48	6.1	666	10	BF628036	BF628036	HVSMF000
C 26	48	6.1	719	9	AV935472	AV935472	AV935472
27	48	6.1	830	10	BF254391	BF254391	HVSMF000
28	48	6.1	844	12	CNS0052P	AL056652	Drosophila
29	48	6.1	1101	12	CNS017SY	AL108860	Drosophila
30	47.8	6.1	307	9	AU070991	AU070991	AU070991
31	47.8	6.1	578	9	AU094712	AU094712	AU094712
32	47.8	6.1	623	9	AU032635	AU032635	AU032635
33	47.6	6.1	517	9	AV933848	AV933848	AV933848
34	47.6	6.1	935	12	CNS006XK	AL066051	Drosophila
35	47.4	6.0	407	10	BE475922	BE475922	946049606
36	47	6.0	657	10	BI717319	BI717319	103101981
37	47	6.0	677	9	AL505169	AL505169	AL505169
38	46.8	6.0	269	10	BE636900	BE636900	WHE1805-1
C 39	46.8	6.0	641	9	AU093820	AU093820	AU093820
C 40	46.8	6.0	669	9	AU096005	AU096005	AU096005
C 41	46.8	6.0	674	9	AU055758	AU055758	AU055758
42	46.6	5.9	691	9	AL504955	AL504955	AL504955
43	46.6	5.9	839	12	CNS004NB	AL054280	Drosophila
C 44	46.4	5.9	453	10	BI779131	BI779131	EBRC01_SO
45	46.4	5.9	557	10	BF618612	BF618612	HVSMF000

## ALIGNMENTS

RESULT 1  
CNS0091P/c  
LOCUS  
DEFINITION  
CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TERT end of BAC #  
BACRI9D16 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION  
AL053013  
VERSION  
AL053013.1 GI:4934461  
KEYWORDS  
GSS.  
SOURCE  
fruit fly.  
ORGANISM  
Drosophila melanogaster

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mosmoser in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

1. 925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"

BASE COUNT	ORIGIN
120 a	61 c
172 t	511 others

Query Match	7.78;	Score 60.8;	DB 12;	Length 925;
Best Local Similarity	13.28;	Pred. No. 0.13;		
Matches	51;	Conservative 180;	Mismatches 155;	Indels 0;
			Gaps	0

QY	251	gcacatcagacgaacgagcaacaacgctggcgccgtatccgcacatctctatccggtgtccac	310
Db	916	SCSBSKSSSSMSTSTSMNSBSCSSCSBSBSSTSTSMSSBSBSBSGSSSSSGTSTACY	857
QY	311	ggggcgccctcgcagagcaatctcggaattgcgcgagcgctgcagagagctgtccgaagtgc	370
Db	856	KCMABSSCGCCGCGMABCMCMSSSSSCGSABARIGKVAASGGAGKRRGGGSGGASMSHS	797
QY	371	gcacgagctgtgacgctcaccgcgcatcaacgaagcgcgccgttgaaatgcgcgtacacatc	430
Db	796	SSACBSSSSSCGASCMASASSSSSSASSRSRSGGAGSGASSRSRSSSSASAGSVSS	737
QY	431	ggttcagtcgccgtaccgaattccgcgcaacgctacacgctcgcgcgccgagctcgatattc	490
Db	736	ASSSSSSCSTSVSCSVASSMSSCSBSBSSSASASSSSSSSSASCAACCSCCTMSCSCT	677
QY	491	ggctgaactgtgaaagcgagcgctgcatgcggcgagcgacgcgcgtgacacgcgtgaaactg	550
Db	676	SASMKAASSSSSSSSSSSSMSSASSSSASASSSSSSGSSSSSGSACGBMSAGGGG	617
QY	551	ccggggccacactgtggaagtctgtglatgcgcgaactgtgactgtgaattacatccctacgcg	610
Db	616	SVSASSGMSSTSVSSGGRSSGSGGGGAGVCGSGSSSBSGSGSGSVCSGSCGCMRCG	557
QY	611	gcacgacgcgccacacacgtctcgatgac	636
Db	556	SSAAAAAASVAAASGMMCGKSGKSGC	531

RESULT	2
CNS0091P	
LOCUS	
DEFINITION	
	CNS0091P 925 bp DNA linear GSS 03-JUN-1991
	Drosophila melanogaster genome survey sequence TE13 end of BAC #
	BA019D16 of Rp11-98 library from Drosophila melanogaster (fruit
	fly), genomic survey sequence.

COMMENT	REFERENCE	AUTHORS	JOURNAL	ORGANISM	SOURCE	KEYWORDS	ACCESSION	VERSION
							AL053013	
							AL053013.1	GI:4934461
						GSS.		
						fruit fly.		
						Drosophila		
						Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
						Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
						Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
						1 (bases 1 to 925)		
						Genoscope.		
						Direct Submission		
						Submitted (02-JUN-1999)		
						Genoscope - Centre National de Sequencage		
						BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
						- web : www.genoscope.cns.fr)		
						Determination of this BAC-end sequence was carried out as part of a		

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammossor in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm)

FEATURES

Location/Qualifiers

1. .925

Source

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1b="RPCI-98"
/clone="BACR19P16"
/note="end : TET3"

```

[illegible]

RESULT	3
LOCUS	CNS00720
DEFINITION	CNS00720 932 bp DNA linear GSS 03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence T7 end of Bac #
VERSION	BAC14809 of Rpt-98 library from Drosophila melanogaster (fruit
KEYWORDS	fly), genomic survey sequence.
SOURCE	AL066742
ORGANISM	AL066742.1 GI:4945205
REFERENCE	GSS.
AUTHORS	fruit fly.
JOURNAL	Drosophila melanogaster.
COMMENT	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 932) Genoscope.
	Direct Submission
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage
	BP 191 91006 Evry cedex - FRANCE (E-mail : segrefigenoscope.cns.fr
	- web : www.genoscope.cns.fr)
	Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoko Oseegawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo.



```

OY      562   tgggaagctcgtatgcacgtagcgacctgatcatatccgtaccgccgcaaggccc    621
DB       375   TTTCAGATTCTGGACACCACGACAGTGTGGAGAAGCACGCACCGACTTGTTCCAACTTGGACCAAGCAGC    434
OY      622   accacgctcgttgaagaagtcttgacctgaagtcattcacatacgaacaacygltccccgcgcg    681
DB       435   GGCTACTTCGCCAGCAGCGGCTCTCGCTTCAAAGGCCGCCACAACCATCATCMCTCGGC    494
OY      682   tacatccggccggaaigtgta    701
DB       495   CACGCCGGGAATCTGCCCCGA    514

RESULT          5
CONS03MW/C     964 bp        DNA         linear    GSS 03-JUN-1999
LOCUS           Drosophila melanogaster genome survey sequence TETJ end of BAC #
DEFINITION      BACR09E09 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION       AL065254
VERSION         AT065254.1 GI:4942606
KEYWORDS        GSS.
SOURCE          fruit fly.
ORGANISM        Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Epiphyarioidea; Drosofilidae; Drosophila.
REFERENCE       Genoscope.
TITLE           Direct Submission
AUTHORS         Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURLNL         BP 191 91006 EVRY cedex - FRANCE (E-mail : seqet@genoscope.cns.fr
COMMENT         - Web : www.genoscope.cns.fr )
                 Determination of this BAC-end sequence was carried out as part of a
                 collaboration with the Berkeley Drosophila Genome Project (BDGP).
                 The BDGP is constructing a physical map of the Drosophila
                 melanogaster genome using these BACs. For further information
                 please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library.html
                 A more detailed description of the library
                 and how to order individual BAC clones, the entire library, or
                 filters for hybridization from the BACPAC Resource Center can be
                 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             source
                     1..964
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone_lib="RPCL-98"
                        /clone-"BACR09E09"
                        /note="end : TETj3"
BASE COUNT       234 a      160 c      60 g      105 t      405 others
ORIGIN
Query Match            6.8%; Score 53.8; DB 12; Length 964;
Best Local Similarity  21.4%; Pred. No. 2.2;
Matches               91; Conservative 154; Mismatches 180; Indels 1; Gaps 1;
```

```

Db      834  AMMSGSSGSGGSSGGSVSRTRTKRRKRSSGGBSGCGKSGSGCGAGGAGGGGGGGRG 775
OY      517  ccggcgcgcgcgcgcgcgttgcacccgttgaactcgtccgcgcgcgcgcacacttgggaatcgtgt 576
Db      774  SSKSSTTSGGSGGGGKRRTTTTKGGSSSGCGTGTSCSGGAAAGSGSGRGVABBGTTGT 715
OY      577  gccgcactgggaactgaattacatgcctcaccgcgcgcgcgcgcacccaccacgtcgcgtgaac 636
Db      714  GSTTSSKVSCKTTTTSKKNYNTSSSYCKSTSTGASNRGTTBSSAAAGSSS-SAAMSMKC 656
OY      637  gagcgtgcactgaagcgccttcatcgaacgcgcgcgtgcgccgcgcgtacatccgcgcgcgcg 696
Db      655  SASVAVGVSCCMCASSSCCSSCMVGGCGKCGTKCGSSSVSRGSSSTRTGCGVCGSSG 596
OY      697  tgcgtctgcacgcggttgcgaagcgccttcgaactcttggaaagcgcgcgcgtctgcga 756
Db      595  SGBTTSSGGGVAANAACASACGCAKGGSSVSGGCGCTTMSRDPDCAAGMACRRVSSGTVA 536
OY      757  agcgcc 762
Db      535  VRBTSR 530

RESULT 6
AC060149
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-047122.R, genomic survey sequence.
ACCESSION AC060149
VERSION AG060149.1 GI:16611379
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoh,Y., Matanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoh,Y., Matanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suehiro-cho,Tsuriumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbases@sc.riken.go.jp, URL:http://hnp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Red process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: MJ3rev
LIBRARY
Vector : pKS145
R.site 1 : SacI
R.site 2 : SacI.
FEATURES
Location/Qualifiers
source
1..700
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-047122.R"
/sex="male"
/cell_type="lymphoblast"
/clone_11b="PTB Chimpanzee Male BAC library"
BASE COUNT 30 a 291 c 334 g 15 t 30 others
ORIGIN
Query Match 6.4%, Score 50.6, DB 12, Length 700;
Best Local Similarity 42.7%; Pred. No. 8.2;
Matches 243; Conservative 0; Mismatches 323; Indels 3; Gaps 1;

```



RESULT	7	
CNS006XK/c		
LOCUS		
DEFINITION		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		

CNS006XK 935 bp DNA linear GSS 03-JUN-1999  
 Drosophila melanogaster genome survey sequence T7 end of BAC #  
 BACR14N09 of RPCR-38 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 AL066051.1 GI:4945019  
 GSS.  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 935)  
 Genome.  
 Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - FRANCE (E-mail : segreifgenoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
 Aaron Mammosser in Pletier de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCR-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library, and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	SOURCE	LOCATION/Qualifiers
	1. .935	
	/organism="Drosophila	melanogaster"
	/db_xref="taxon:7227"	
	/clone_1lb="RRC1.98"	
	/clone="BACRI4N09"	
	/note="end : 47"	
BASE COUNT	257 a 170 c 162 g	96 t 250 others
ORIGIN		

Query Match	6.4%	Score 50.4	DB 12	Length 935
Best Local Similarity	29.0%	Pred. No. 9		
Matches 123	Conservative 110	Mismatches 184	Indels 7	Gaps 2

[illegible]

RESULT	8	
LOCUS	BE040131/c	
DEFINITION	BE040131	1022 bp mRNA linear EST 07-JUN-2000
ACCESSION	OD1020208 OD Oryza sativa cDNA 5'	transcription factors, mRNA sequence.
VERSION	BE040131	
KEYWORDS	BE040131.1 GI:8335147	EST.
SOURCE	Oryza sativa.	
ORGANISM	Oryza sativa.	
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Eriatoidaeae; Oryzeae; Oryza. 1 (bases 1 to 1022) Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferreira H.N., Kawasaki,S., McColllough,A., Michalowski,C.B., Palciado,C., Scara,G., Wheeler,M. and Zepeda,C.R.: Functional genomics of Plant Stress Tolerance Unpublished (2000)	
TITLE	JOURNAL	Contact: Michalowski,C.B.
COMMENT		University Of Arizona



Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 1009)  
 Genoscope.  
 AUTHORS  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segreif@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.  
 FEATURES  
 Location/Qualifiers  
 1..1009  
 /organism="Drosophila melanogaster"  
 /plasmid="pBelobAC11"  
 /db\_xref="taxon:7227"  
 /clone\_lib="DrosBAC"  
 /clone="BACN03p19"  
 /note="end : Sp6"  
 BASE COUNT 147 a 377 c 178 g 64 t 243 others  
 ORIGIN  
 Query Match 6.4%; Score 50; DB 12; Length 1009;  
 Best Local Similarity 31.0%; Pred. No. 11; Mismatches 115; Conservative 83; Mismatches 172; Indels 1; Gaps 1;  
 Oy 396 caagacggcgctgagatgcgcgctacagacatggttcagtcgcgtcaagatccg 455  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 1006 CCRBRSSAIVGMAVVCVCGSARCGAASCMMACAGCGCMRARGAIVTCCGG 947  
 Oy 456 caagcgtacagcg-gcgcgccgagctgagtctgctgaactggaacggcgcg 514  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 946 MGCCGGGMSAGRCRSCCGGGGSCSCSSCCSSCGGCGSSSGGCGSSGNGMSG 887  
 Oy 515 tgcggcgccgacggcgctggtgcaacgctggaacggcgccactgggaagtcg 574  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 886 KGSRSKCCCGSGSGGGGCGCCSCCMGRCGGGGCGACCGSSGSSGSSGS 827  
 Oy 575 atgcgactggaactgaattacatgcctacacggcgcaagcggccacacgtcg 634  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 826 SGSSGSSCGCGCGCGCGCGGGGSGSSGCGCACGCGCNSGSSGSGGCGGS 767  
 Oy 635 gcgaagctgaactgaagccttcacgcgcggtcgccgcggtacatccgcg 694  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 766 SSSGSSGSSGSSGSGGRCGCGCGCGCGGCGGCGCCGCCSCCSCSSCGCC 707  
 Oy 695 agtgcgtatcgatggtgagagcggttcgaactctggaggggcggtctgc 754  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 706 GGCRCRSCVCGCGGGGGGGGGGSGKGMCGCGCGGSGSGGGGSGBCCCGS 647  
 Oy 755 gaagcgccgat 765  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 646 GGGCCCCCSMY 636  
 RESULT 11  
 BE517742 449 bp mRNA linear EST 08-AUG-2000  
 LOCUS WHE0802\_C12\_E24Z5 Wheat vernalized crown cDNA library Triticum  
 DEFINITION aestivum cDNA clone WHE0802\_C12\_E24, mRNA sequence.  
 ACCESSION BE517742  
 VERSION BE517742.1 GI:9741772  
 KEYWORDS EST.  
 SOURCE bread wheat.  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Triticum.  
 1 (bases 1 to 449)  
 Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han ,P.S., Hsie,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Jong,J.C.  
 The structure and function of the expressed portion of the wheat genomes - Vernalized crown cDNA library  
 Unpublished (2000)  
 Contact: Olin Anderson  
 US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 5105595773  
 Fax: 5105595818  
 Email: oanderson@w.usda.gov  
 Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
 Seq primer: StrataGene SK primer.  
 Location/Qualifiers  
 1..449  
 /organism="Triticum aestivum"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"  
 /clone="WHE0802\_C12\_E24"  
 /clone\_lib="Wheat vernalized crown cDNA library"  
 /tissue\_type="Crown tissue of seedling"  
 /dev\_stage="Five-week old seedling"  
 /lab\_host="E. coli SOLR"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site\_1: EcoRI; Site\_2: XhoI; Seeds were germinated and grown at 4 C for 5 weeks. The tissue, total RNA, and poly(A) RNA were prepared. A cDNA library was made, and the cDNA clones were in vivo excised to give plasmid phagemide in the TJ Clonase lab (Choi, Close, Fenton) at the University of California, Riverside. Reverse transcribed DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."  
 BASE COUNT 63 a 175 c 137 g 74 t  
 ORIGIN  
 Query Match 6.3%; Score 49.8; DB 10; Length 449;  
 Best Local Similarity 47.8%; Pred. No. 11; Mismatches 144; Conservative 0; Mismatches 157; Indels 0; Gaps 0;  
 Oy 387 caacgcgcatcagcagcgccgctgtaatgcccctacgacatctgttcacgtcac 446  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 128 CAGCGCGGTGGCCAGAGCCACGCGCGGCCACCTAACGGGTGTTCAGCGCTTCTG 187  
 Oy 447 gaattccgcaacgagctacagcgcgcgagctgatatctgtgctgaactgaagcg 506  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 188 GGAACGGGCCAACCTCTCCCGCGCGCGCGCATCAACACGGCGGACCCGAAG 247  
 Oy 507 cggcggtatgcggcgcgacgctgtgacacgctggaactgacggcgagcgag 566  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 248 GAGCGTCATGCGCGGCTCGCGCGCGCACAGCTGTGACATGTCAGAGGTCTCTCCG 307  
 Oy 567 agtcgtgatcgactggagcgaactacatcgctacgcgcgcaagcggccacac 626  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 308 GCGGCGCTCCGTGATCTGTTGGTGGCCAAAGCCGTGGCTCTCGGCTTCATCA 367  
 Oy 627 gtccgtgagcgaactggaacgcttcacgcgacgacgagcggtgcggcgctacat 686  
 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
 Db 368 CGAGTACGGGCTGACAGGGGTGACGTCGATCAAGAGCGCTTGCGCGCGCTCTCCG 427  
 Oy 687 c 687  
 |  
 Db 428 c 428  
 RESULT 12  
 A0093968/c

LOCUS	AU095988	682 bp	mRNA	linear	EST 30-JUN-2000
DEFINITION	AU095988 Rice green shoot Oryza sativa cDNA clone S11827,				mRNA
ACCESSION	AU095988				
VERSION	AU095988.1	GI:8858670			
KEYWORDS	EST.				
SOURCE	Oryza sativa.				
ORGANISM	Oryza sativa				
REFERENCE	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartioideae; Oryzaceae; Oryza. 1 (bases 1 to 682) Sasaki,T. and Yamamoto,K. Rice cDNA from green shoot (2000) Unpublished (2000) Contact: Takuji Sasaki National Institute of Agrobiological Resources Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan Tel.: 81-298-38-7441 Fax: 81-298-38-7468 Email: tsasakia@agr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/ PROJECT "RGF". S11827_102.				
FEATURES	Location/Qualifiers				
source	1..682 /organism="Oryza sativa" /strain="Nipponbare" /db_xref="taxon:4530" /clone="S11827" /clone_1fb="Rice green shoot" /note="Green shoot (8 days old)" /project="RGF".				
BASE COUNT	153 A 184 C 212 G 132 T			1 others	
ORIGIN					
Query Match	6.3% Score 49.6; DB 9; Length 682;				
Best Local Similarity	47.1%; Pred. No. 12;				
Matches	185: Conservative 0; Mismatches 205; Indels 3; Gaps 1;				
Dy	299 tcggagtcgcaactcgggcgtcgacgaagaattccggatctccgcgcgcgcgtctgacaagc	358			
Dd	453 tggccttccgcgagagctgggtctcgaaacgacgcgagctgtttccagctccagtaca	394			
Dy	359 tctcgagcgtgcgcgcgcgtctgaaagctcacgcgatcacgaacgcgcgcgttgatgccg	418			
Dd	393 gcccgacctgtcagctgcggctgcggccaccacacctgcggccgcttccggactctccg	334			
Dy	419 ccatacgaactcgtgtcagctccgcgtcaagaattccgcgaacgtcacgaacgcgcgcgcg	478			
Dd	333 ccttcgacacgaacccccttgcgtcagaaacgcttggaacgcgcgggaacacgacccgaa	274			
Dy	479 agctgcatcttgcttgaacttgaaacgcgcgtgatgcgcgcgcgcgcgcgcgcgcgcgc	538			
Dd	273 gcgggatgtgcctgcggaacaaacccctgcgtgcgcatgcaccgcggtaaacacgacgga	214			
Dy	539 ccgtggaactcgc	598			
Dd	213 acgtatccccgtct---ccttgtgaaattctcgggtgacattcctgcattcgtcacatgt	157			
Dy	599 tcgacctaccgc	658			
Dd	156 tacagatctgtctgctgttcgttatcgccttgaggccgcctcgaagcttcgcggcgctgatca	97			
Dy	659 tcgacgaacgc	691			
Dd	96 tcgcccacccgttctgcggccgcgacnccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	64			
RESULT 13					
AZ934730/c	AZ934730	534 bp	DNA	linear	GSS 24-APR-2001
LOCUS	BJ_Ba00002f2ar B. japonicum BAC library Bradyrhizobium japonicum				
DEFINITION					

[illegible]



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 12:44:44 ; Search time 6219.51 Seconds

(without alignments)  
2462.931 Million cell updates/sec

Title: us-10-003-759-3\_COPY\_52\_783

Perfect score: 732  
Sequence: 1 tgcagctgctcttcccca.....atttccgaacggtcag 732

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_da:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pl:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hcgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
------------	-------	-------	-------	--------	-------	-------------

1	732	100.0	786	6	AX339682	AX339682 Sequence
2	732	100.0	1877	6	AX339680	AX339680 Sequence
3	692.8	94.6	1874	1	RM072637	U72637 Rhodothermu
4	131.8	18.0	1116	1	AF233376	AF233376 Streptomy
5	131.8	18.0	1116	6	AR129926	AR129926 Sequence
6	131.8	18.0	1116	6	AR168360	AR168360 Sequence
7	105.8	14.5	1716	6	AR168362	AR168362 Sequence
8	97.8	13.4	3527	1	SHU51222	U51222 Streptomyce
9	96.2	13.1	1470	1	SREGLS	X73953 S.rochei eg
10	89	12.2	1311	1	AF130408	AF130408 Streptomy
11	87.6	12.0	1565	1	SLH04629	U04629 Streptomyce
12	87.6	12.0	41782	1	SCG11A	AL133210 Streptomy
13	69	9.4	11202	1	AE005123	AE005123 Halobacte
14	68.8	9.4	4262	1	AF335723	AF335723 Burkholde
15	63.2	8.6	14483	1	AE006992	AE006992 Mycobacte
16	63.2	8.6	67200	1	MTV017	AL021897 Mycobacte
17	57.6	7.9	1983	6	AX047403	AX047403 Sequence
18	55	7.5	36368	1	SC9B5	AL035206 Streptomy
19	54.8	7.5	346897	1	AP002995	AP002995 Mesorhizo
20	54.6	7.5	45313	1	SCD95A	AL357432 Streptomy
21	54.6	7.5	125020	9	AF429315	AF429315 Homo sapi
22	54.4	7.4	38995	1	SCF34	AL109974 Streptomy
23	54	7.4	34182	1	SC111	AL096849 Streptomy
24	53.6	7.3	77534	1	AP235504	AP235504 Streptomy
25	53	7.2	33779	1	SCGD3	AL096822 Streptomy
26	52.2	7.1	1849	1	SGSTRB1F	X78973 S.galbus DS
27	52	7.1	349116	1	AP003003	AP003003 Mesorhizo
28	51.4	7.0	3644	5	AF034576	AF034576 Gallus ga
29	51.4	7.0	125020	9	AF429315	AF429315 Homo sapi
30	51.2	7.0	22115	1	SCC22	AL096839 Streptomy
31	50.6	6.9	3525	1	SGSTRB1FG	AL78974 S.glaucisce
32	50.6	6.9	14596	1	AE007168	AE007168 Mycobacte
33	50.6	6.9	25459	1	SCGA6985	AL006985 Streptomy
34	50.6	6.9	30302	1	MTCY66G11	292774 Mycobacteri
35	50.6	6.9	40806	1	MSGY23	AD000016 Mycobacte
36	50	6.8	29625	1	SC422A	AL159178 Streptomy
37	50	6.8	36394	1	SCD35	AL160312 Streptomy
38	50	6.8	40104	1	SCF11	AL132662 Streptomy
39	49.8	6.8	36734	1	SCG610	AL049497 Streptomy
40	49.6	6.8	26555	1	SC2G1	AL391014 Streptomy
41	49.6	6.8	146555	2	OSJN00013	AL606448 Oryza sat
42	49.6	6.8	215050	1	AL646057	AL646057 Ralstonia
43	49.4	6.7	150814	2	AP003541	AP003541 Oryza sat
44	49.2	6.7	12130	1	AE005133	AE005133 Halobacte
45	49.2	6.7	41173	1	SC141	AL132648 Streptomy

#### ALIGNMENTS

RESULT	1	AX339682	786 bp	DNA	Linear	PAT 10-JAN-2002
LOCUS	AX339682	Sequence 3 from Patent WO0196382.				
DEFINITION	AX339682					
ACCESSION	AX339682.1	GI:18135684				
VERSION	AX339682.1	GI:18135684				
KEYWORDS						
SOURCE						
ORGANISM		Rhodothermus marinus.				
		Rhodothermus marinus				
REFERENCE		Bacteria; CFB group; Rhodothermus group; Rhodothermus.				
AUTHORS		1 (sites)				
		Wicher,K.B., Holst,O.P., Hachem,M.Y., Karlsson,E.M. and				
		Hreggvidsson,G.O.				
TITLE		Thermostable cellulase				
JOURNAL		Patent: WO 0196382-A 3 20-DEC-2001;				
		Prokaria ehf. (IS)				
FEATURES		Location/Qualifiers				
source		1..786				
		/organism="Rhodothermus marinus"				
		/db_xref="taxon:29549"				
BASE COUNT		141 a 236 c 279 g 130 t				
ORIGIN						

Query Match 100.0%; Score 732; DB 6; Length 786;  
Best Local Similarity 100.0%; Pred. No. 4,7e-95;  
Matches 732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 tgcgaactgctcttcccgatgacgaacgaagaaacgcctgaagcccgagcg 60
    |||
DB 52 TCGGACGTGGCTTTCCTCCGATGGCGAACAAGGAAGAACCGGAGCCCTGAGCCGAGCCG 111
    |||

QY 61 accgtcgaactgtgctggaacgtcggagcgccgagatgtgcccggggggcctaccgggtg 120
    |||
DB 112 ACCGTGAGCTGTGGCGACCTGGAGCGCGCGGATGTGGCCGGGGGGCCCTACCGGGTG 171
    |||

QY 121 atcaacaacgtatgagggcgcgagaccccgccatgtcaatgaggtgcgactggaacagggc 180
    |||
DB 172 ATCAACAACGTATGGGGCGCGGAGACCGCCGACGATGAGGTGCGGACTGGAAGCGGGC 231
    |||

QY 181 aacttcacgatacacaagggccgatacagaacaacgcaacagctgacgctatccggcc 240
    |||
DB 232 AACCTACAGATCACAGGGGCGGATCACGACAAAGCAACGTGGCCGCTATCCGGCC 291
    |||

QY 241 atctactcgggtgctcgaactggcgccctgacgagcaattccggatgtccggcgcggtg 300
    |||
DB 292 ATCTACTCGGGTGGCTCCACTGGGGCGCTGCACAGCAATTCCGGATTTGCCGGCGGTG 351
    |||

QY 301 caggagactgtcgcagctgacgacgagctggaacgtcgaacgcaacgacggcgcgcttg 360
    |||
DB 352 CAGGAGCTGTCCGACGCGCCGACGAGCTGGACGCTCACGCCGATCACAGCGGGCCGCTG 411
    |||

QY 361 aatgacgacgaacgaactgtgttcagctccgtaacgaattccggcaacggcctacagggc 420
    |||
DB 412 AATGCGCGCTACGACATCTGGTTAGTCCGCTACAGATTCGCGAAGGCTACAGCGGC 471
    |||

QY 421 gggcgacgactgatactgctgcaacttggaacgagcgcgctgatacgcgcgacggcg 480
    |||
DB 472 GGGCGCGAGCTGATGATCTGGCTGMACTGGAAGCGCGCTGATGCCGGCGGCGAGCGCC 531
    |||

QY 481 gtggcaccgttggaactgagcgagggccacactgggaagctgtgatacgcaactggagctg 540
    |||
DB 532 GTGGCACCCTGGAGACTGGCGGGGGCCACCTGGGAAGTGTGATGCGGACTGGAGCTGG 591
    |||

QY 541 aattacatgcctacacggcgacgaagcccaacgctcgtgtgagcgagctggagccggaag 600
    |||
DB 592 AATTACATCTCCCTAACCGGCGCACGAGCCACACAGTGGTGAGCGAGCTGGAACGCTGAAG 651
    |||

QY 601 gacctacgaacgagcggtgcgcgcggcggtacatccggcgcgagtgatgtcgcagtg 660
    |||
DB 652 GCCTTCATCGACAGCGGGTGGCGCGGCGGTACATCCGGCGGAGTGATCTGCAATGCG 711
    |||

QY 661 gtggagacgggcttcgaactcgtggaagggcgggcggtcgtggaagcgcgcaattttcc 720
    |||
DB 712 GTGGAGACGGGCTTCGAACCTGGGAGGGGGGCGGCTGTGCAAGCGCGCATTTTTC 771
    |||

QY 721 gtaacggtgacg 732
    |||
DB 772 GTAACGGTGCAG 783
    |||
```

## RESULT 2

AX339680 1877 bp DNA linear PAT 10-JAN-2002  
LOCUS AX339680  
DEFINITION Sequence 1 from Patent WO0196382.  
ACCESSION AX339680  
VERSION AX339680.1 GI:18135682  
KEYWORDS  
SOURCE  
ORGANISM Rhodothermus marinus.  
Rhodothermus marinus  
Bacteria; CFB group; Rhodothermus group; Rhodothermus.  
REFERENCE  
1 (sites)  
AUTHORS Wicher, K.B., Holst, O.P., Hachem, M.Y., Karlsson, E.M. and  
Hreggvidsson, G.O.  
TITLE Thermotable cellulase  
JOURNAL Patent: WO 0196382-A 1 20-DEC-2001;

Prokaria ehf. (15)  
FEATURES  
source  
Location/Qualifiers  
1..1877  
/organism="Rhodothermus marinus"  
/db\_xref="taxon:29549"  
710..1495  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAD20684.1"  
/db\_xref="GI:18135683"

## CDS

WDARVAGGRYRVINNWGAETACIEVLEGTGNTITRADHNGNNAAIPATIEGC  
HMGACTSNGLPFRVQELSDVPTSWTLPITGRRNAAVDIWFSPVTSNGQYSGAE  
LMIWLNMGVMPDGSRYATVELAGATVEWADMDMYIAYRRPTPTSVSELDLKA  
FIDDAVARVIRPEWYLNHAVETGFELMEGAGLRFADRSVTVQ"

BASE COUNT 331 a 582 c 632 g 332 t

## ORIGIN

Query Match 100.0%; Score 732; DB 6; Length 1877;  
Best Local Similarity 100.0%; Pred. No. 3,7e-95;  
Matches 732; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 tgcgaactgctcttcccgatgacgaacgaagaaacgcctgaagcccgagcg 60
    |||
DB 761 TCGGACGTGGCTTTCCTCCGATGGCGAACAAGGAAGAACCGGAGCCCTGAGCCGAGCCG 820
    |||

QY 61 accgtcgaactgtgctggaacgtcggagcgccgagatgtgcccggggggcctaccgggtg 120
    |||
DB 821 ACCGTGAGCTGTGGCGACCTGGAGCGCGCGGATGTGGCCGGGGGGCCCTACCGGGTG 880
    |||

QY 121 atcaacaacgtatgagggcgcgagaccccgccatgtcaatgaggtgcgactggaacagggc 180
    |||
DB 881 ATCAACAACGTATGGGGCGCGGAGACCGCCGACGATGAGGTGCGGACTGGAAGCGGGC 940
    |||

QY 181 aacttcacgatacacaagggccgatacagaacaacgcaacagctgacgctatccggcc 240
    |||
DB 941 AACCTACAGATCACAGGGGCGGATCACGACAAAGCAACGTGGCCGCTATCCGGCC 1000
    |||

QY 241 atctactcgggtgctcgaactggcgccctgacgagcaattccggatgtccggcgcggtg 300
    |||
DB 1001 ATCTACTCGGGTGGCTCCACTGGGGCGCTGCACAGCAATTCCGGATTTGCCGGCGGTG 1060
    |||

QY 301 caggagactgtcgcagctgacgacgagctggaacgtcgaacgcaacgacggcgcgcttg 360
    |||
DB 1061 CAGGAGCTGTCCGACGCGCGCACGAGCTGGAGCGCTCACGCCGATCACAGCGGGCCGCTG 1120
    |||

QY 361 aatgacgacgaacgaactgtgttcagctccgtaacgaattccggcaacggcctacagggc 420
    |||
DB 1121 AATGCGCGCTACGACATCTGGTTAGTCCGCTACAGATTCGCGAAGGCTGACAGCGGC 1180
    |||

QY 421 gggcgacgactgatactgctgcaacttggaacgagggcggtgatacgcgcgacggcg 480
    |||
DB 1181 GGGCGCGAGCTGATGATCTGGCTGMACTGGAAGCGCGCTGATGCCGGCGGCGAGCGCC 1240
    |||

QY 481 gtggcaccgttggaactgagcgagggccacactgggaagctgtgatacgcaactggagctg 540
    |||
DB 1241 GTGGCACCCTGGAGACTGGCGGGGGCCACCTGGGAAGTGTGATGCGGACTGGAGCTGG 1300
    |||

QY 541 aattacatgcctacacggcgacgaagcccaacgctcgtgtgagcgagctggagccggaag 600
    |||
DB 1301 AATTACATCTCCCTAACCGGCGCACGAGCCACACAGTGGTGAGCGAGCTGGAAG 1360
    |||

QY 601 gacctacgaacgagcggtgcgcgcggcggtacatccggcgcgagtgatgtcgcagtg 660
    |||
DB 1361 GCCTTCATCGACAGCGGGTGGCGCGGCGGTACATCCGGCGGAGTGATCTGCAATGCG 1420
    |||

QY 661 gtggagacgggcttcgaactcgtggaagggcgggcggtcgtggaagcgcgcaattttcc 720
    |||
DB 1421 GTGGAGACGGGCTTCGAACCTGGGAGGGGGGCGGCTGTGCAAGCGCGCATTTTTC 1480
    |||

QY 721 gtaacggtgacg 732
```



|||||  
Db 1481 GTAACGGTGCAG 1492  
RESULT 3  
LOCUS RMU72637 1874 bp DNA linear BCT 11-MAY-1999  
DEFINITION Rhodothermus marinus cellulase (celA) gene, complete cds.  
ACCESSION U72637  
VERSION 072637.1 GI:2304960  
KEYWORDS  
SOURCE Rhodothermus marinus.  
ORGANISM Rhodothermus marinus.  
REFERENCE 1 (bases 1 to 1874)  
AUTHORS Halldorsdottir, S., Thorolfsson, E.T., Spilliaert, R.,  
Johannsson, M., Thorbjarnardottir, S.H., Palsdottir, A.,  
Heggvissdottir, G.O., Kristjansson, J.K., Holst, O. and Eggertsson, G.  
Cloning, sequencing and overexpression of a Rhodothermus marinus  
gene encoding a thermostable cellulase of glycosyl hydrolase family  
12  
JOURNAL Appl. Microbiol. Biotechnol. 49 (3), 277-284 (1998)  
MEDLINE 98242392  
REFERENCE 2 (bases 1 to 1874)  
AUTHORS Halldorsdottir, S.  
TITLE Direct Submission  
JOURNAL Submitted (25-SEP-1996) Department of Molecular Biology,  
Institution of Biology, Grensasvegur 12, Reykjavik 108, Iceland  
FEATURES  
source  
1. 1874  
/organism="Rhodothermus marinus"  
/strain="TT1378"  
/db\_xref="taxon:29549"  
710..1493  
/gene="celA"  
710..1492  
/gene="celA"  
/EC\_number="3.2.1.4"  
/note="belongs to family H of cellulase catalytic domains;  
endo-1,4-beta-glucanase"  
/codon\_start=1  
/transl\_table=11  
/product="cellulase"  
/protein\_id="AA055594.1"  
/db\_xref="GI:2304961"  
/translation="MNVRAVLVLSLLFSCDMLFPDGDNGKEPEPEPEPTVELGCR  
WDARDAVAGRYRYINNVWGAETACIEVGETENFTTTRADHNGNNVAAPATYGC  
HMAPARAIRPCARAGAVRAHELDVPTITGDMANAYDIWSPVNSGMSGAGEL  
MIMLNNGCMVPGSGSRVATVELAGATMEVYVADMDNRYIARRTTPTTSVELDLAF  
IDDAVARGTIRPEWTLHAYETGELMEGGAGLRTADPSYVQ"  
BASE COUNT 331 a 580 c 631 g 332 t  
ORIGIN  
Query Match 94.6% Score 692.8; DB 1; Length 1874;  
Best Local Similarity 99.3%; Pred. No. 1.4e-89;  
Matches 721; Conservative 0; Mismatches 2; Indels 3; Gaps 3;  
QY 1 tgcagctgctcttcccgatgacgacgaaagaaacggagcctgagcccgagcgcg 60  
Db 761 TGCAGCGGCTCTTCCGATGGCGACACGAAAGAACCGGAGCCTGAGCCCGAGCCG 820  
QY 61 acgcgtgagctgtgagagcgttgagacgcgcgcgcatgtgcccggggggcgctaccgggtg 120  
Db 821 ACCCTCGAGCGTGTCCGACGCTGGAGCGCGCGCATGTGGCGGGGGCGCTACCGGGTG 880  
QY 121 atcaacaacratatggggcgagagaccccgagtcattgagtcgagacttgaaacgggc 180  
Db 881 ATCAACAACGATGGGGCGGAGACCGCCAGTGCATGTGAGTGGACTGGAACGGGC 940  
QY 181 aacttcagatcacagggcgacacacgacacgacgacgacgacgacgacgacgacgacgacgac 240  
Db 941 AACTTCAGATCACAGGGCGGATCAGACACAGGCAACACGTCGCGCTATCCGGCC 1000

QY 241 atctacttcgggtgacgctgagggcgccctgacagacgaattcgaggtccggcgctg 300  
Db 1001 ATCTACTTCGGGTGCCACT-GGGCGCCTGCACGACGAAATTCGGGATTTG-CGGCGCGCTG 1058  
QY 301 caggagctgtccgagcgtgagcgacagagcttgagactcaacgcgcgttaacgaaggccgctg 360  
Db 1059 CAGGAGCTGTCCAGACGTCGCGACGAGCTGGACG-TCACGCCGATCACGAGCGGCGCCTGG 1117  
QY 361 aatgcgcgacacacatctgttcagtcgccgtaacgacattccggcaacggtctaacgcgc 420  
Db 1118 AATGCCCTTACGACATCTGTTCAGTCCGCTCAGCAATTCGGCAACGGCTTACAGCGC 1177  
QY 421 ggcgcgagctgatgatctgctgctgacgtgaaagcgagcggtgacgtgagcgagcgcg 480  
Db 1178 GCGCGCAGCTGATGATCTGGTGAATCTGGAACTGGAAAGCGCGCGCTGATGCGCGGCGCAGCCG 1237  
QY 481 gtggccacgctgagacgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 540  
Db 1238 GTGCCACCGCTGGAACTGGCGGGGCCACCTGGGAACTGTGTATGCCGACTGGACTGG 1297  
QY 541 aattacatgcctacacgagcgacagacccaccacgctcgtgtagcgagcgagcctgaa 600  
Db 1298 AATTACATCGCCTTACCGCGCGACGACGCCACACAGTCTGAGCGAGCTGGAAC 1357  
QY 601 gcttcacgacagcagcgcgtgc 660  
Db 1358 GCCTTCATTCAGACGACGCGGGTCCGCCCGCGGCTACATCCGCGGAGTGTATCTGCAATGCG 1417  
QY 661 gtggaagacgagcttgcgaactctggagggcgagcgagcgagctgagcgagcgagcgttttc 720  
Db 1418 GTGGAAGACGGGCTTTCGAATCTGGAGAGCGGGCGCGCTGTGGAAAGCGCGCATTTTTC 1477  
QY 721 gtaacggtgcag 732  
Db 1478 GTAACGGTGCAG 1489  
RESULT 4  
LOCUS AF233376 1116 bp DNA linear BCT 02-AUG-2000  
DEFINITION Streptomyces sp. 11A68 cellulase 12A (cell12a) gene, complete cds.  
ACCESSION AF233376  
VERSION AF233376.1 GI:9651812  
KEYWORDS  
SOURCE Streptomyces sp. 11A68.  
ORGANISM Streptomyces sp. 11A68  
REFERENCE 1 (bases 1 to 1116)  
AUTHORS van Solingen, P., Meijer, D., van der Kleij, W.A.H., Barnett, C.C.,  
Bolle, R., Power, S.D. and Jones, B.E.  
Cloning and expression of an endocellulase gene from a novel  
Streptomyces isolate from an East African soda lake  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 1116)  
REFERENCE van Solingen, P., Meijer, D., van der Kleij, W.A.H., Barnett, C.C.,  
Bolle, R., Power, S.D. and Jones, B.E.  
Direct Submission  
TITLE Submitted (10-FEB-2000) Microbial & Molecular Screening, Genencor  
JOURNAL International B.V., Archimedesweg 30, 2333 CN Leiden, The  
Netherlands  
FEATURES  
source  
1. 1116  
/organism="Streptomyces sp. 11A68"  
/strain="11A68"  
/db\_xref="taxon:133452"  
1..1116  
/gene="cel12A"  
1..93  
misc\_feature  
/gene="cel12A"  
/note="Region: signal sequence"  
1..1116  
CDS



QY	647	ggtatctcgtacgttggagagcccttcgaacitctggagagcgagcgccggtctgcgaa	706
Db	671	GGTACTCTACCAAGCATTCAGGCGGCGCTTCGAGCCTGGGAGGGCGGACACGGCTTGCCG	730
QY	707	gcgcgcgaattttccgttaacggtgca	721
Db	731	TGAACCTCCTCTCCTCCGCGGTGAA	755
RESULT	6		
LOCUS	AR168360	1116 bp	DNA
DEFINITION	Sequence 2 from patent US 6287839.		linear
ACCESSION	AR168360		PAT 17-DEC-2001
VERSION	AR168360.1	GI:17904235	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1116)		
AUTHORS	Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Meyler,W.		
TITLE	Cellulase producing actinomycetes, cellulase produced therefrom and method of producing same		
JOURNAL	Patent: US 6287839-A 2 11-SEP-2001;		
FEATURES	Location/Qualifiers		
SOURCE	1..1116		
BASE COUNT	191 a 432 c 340 g	153 t	
ORIGIN	/organism="unknown"		

Query Match	18.0%:	Score 131.8:	DB 6:	Length 1116:
Best Local Similarity	50.7%:	Pred. 6.9e-107:		
Matches 347:	Conservative	0:	Mismatches 332:	Indels 6:
				Gaps 1:
OY	47	ctgaagcccgagccgaacgcgctcgaagctcgtgcgagcgtctggaacgcgcgcgaatcgtgcgcggg	106	
Db	77	CGGGTCCCGCCCGACGCGAACCAGCAGATCTCGGACCGCTACGGGCACCCACAGATCCAG	136	
OY	107	ggcgctaccgggtgatacaacaafatgggggggggagacggcccaagtcattgagtcg	166	
Db	137	ACCGGTACGTGTTGCGAGAACACACCCCTGGGGACACGCGCACCCAGTGCATCAATGTGA	196	
OY	167	gacttgaaacgggcgaacttcacgatalcaacacgggcgcatcagacaacgcgaacgcttg	226	
Db	197	CCGGCAAGGTTTCGAGATCACACCAGGCGGAGAGTTCGGTGCAGCAACAGCGGCCCCGA	256	
OY	227	ccgcatcaccggcccatctcaacttcggtgctcgaatggggcgctcgcacgagcaattcgggat	286	
Db	257	AGTCTATTCCTCGGTCTACGACGGCTCCCATACGGCAACTGCCTGCCGCCCCACAGACGC	316	
OY	287	tgcgcgcgcgctgcaagagagctcgtcgcagcgtgcacgagctcgagcgtcaccgcgatca	346	
Db	317	TGCCCATCGCGATCAGTCTCGATCGGACGGGCCCCAGCAGTGTCTTCATCCGCTACACCG	376	
OY	347	cgaacggccgcttgaaatgcgcgcctacgacatcttggttcaagttccgctcgaattccgcga	406	
Db	377	GCAACGGGCTCTACACCCCGGTACGCATCTGG-----CTGAACCCGACACCCCGGA	430	
OY	407	acggctacaagcgcgcgcgcgcagctcgaatcgtatctgactgaatgaaacgcgcgcgtgatgc	466	
Db	431	CCAACGGGTGTAACCGGACCGGACGAGTATGATCTGTTCAACCGGGTGGCCCCGGTCCAGC	490	
OY	467	cgagcgcgacgcgcgctgctgcacacgtgaaactcgcgcgggccaactcgggaagtctgatatg	526	
Db	491	CCATTCGGTTCGCGCGTCCGACAGGCCACAGTCTGGGGGCGGCAAGCTTGGAGGTGTGACCG	550	
OY	527	ccgaactggacttggaatatacgcctcaccgcgcgcacgacgcacacacgcgtctgcgtacgcg	586	
Db	551	GCAGCAACGTTTCGAACGACGTGATCTCTTCCCTGGCCGCCCTCCGCGATACACACTGGA	610	
OY	587	agctggacctgaagccttcatacgacgcgcggtccgcgcgactacatccggccggagt	646	

Db	611	GCTTCGACGTGAAGGACTTGTCTGCAGCAGGGCCGTAGCCACGGCCTTG6CACAACCCGGGACT	670
Oy	647	ggtatctgcagatcgagtcgagacacgggcttcgaactctggagagggcgggccggtctcgaa	706
Db	671	GGTACCTTACCAGCAGCATCCAGGCGGGCTTCGAGCCTGTGGAGGGCGGCACCGGTCTGGCCG	730
Oy	707	gcgcgcgattcttcgatacggctgaa	731
Db	731	TGAACTCCTTCTCTCCGCCGGGTGAA	755
RESULT	7		
LOCUS	ARI68362	1716 bp	DNA
DEFINITION	Sequence 4 from patent US 6287839.		linear
VERSION	ARI68362		PAT 17-DEC-2001
KEYWORDS	ARI68362.1 GI:17904238		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 1716)		
TITLE	Jones,B.E., Van Der Kleij,W.A.H., Van Solingen,P. and Weyler,W.		
JOURNAL	Cellulase producing actinomycetes, cellulase produced therefrom and		
FEATURES	method of producing same		
source	Patent: US 6287839-A 4 11-SEP-2001;		
	Location/Qualifiers		
	1..1716		
	/organism="unknown"		
BASE COUNT	307 a 611 c 559 g	239 t	
ORIGIN			

Query Match	14.5%;	Score 105.8;	DB 6;	Length 1716;
Best Local Similarity	53.5%;	Pred. No. 3e-06;		
Matches 269;	Conservative 0;	Mismatches 22;	Indels 7;	Gaps 2;
QY 230	ccatcaccgacatctactctcgtggtgcacatcgtggcgctgcacagacgaattcggatctgc	289		
Db 711	cctatccctcggtggttacgacggcgtccactaagccaaactgcgcgccgacagcgttcg	770		
QY 290	cgcggcgctgcaggaagctctgtccgaagctgtgcagagctctgaaagctcaagccgatcaaga	349		
Db 771	ccatccgggatcagctcgtcgtacggcaccggccacagcagtgcttcacacgcctaacaccggca	830		
QY 350	cggcgccgcttgtaatgcgcgcctacagacatctggtcagctccggtcaagaattccggacag	409		
Db 831	acggcgtttacacgcccccgttacgacatcttgctt-----gaccgcacaccccccaccca	884		
QY 410	gctaaagcggcgcgccgcagcttgatgatactgctgaacttgaaacggcgcgcttgatgcgg	469		
Db 885	acggcggttaacccggacggacgagatcagatctggttcaaccgggtgcgccgggtccagccca	944		
QY 470	gcgcgacgcgcgttgccacgcgttgaaactgtgccggggccaaactgggaagtctgtatgcg	529		
Db 945	tccggttcccccgtccggcaccggccaccgctccggcgccgacagctggagagtgatggacggca	1004		
QY 530	actgggaactggaattacatcgctcctacccggcgccagacgcccacacccgcgtgtagcgggc	589		
Db 1005	gcAACGGTTCGAACAGCAGCTGATCTCTCTCTGGCCCCCTCCGGCATGACAGACTGGAGCT	1064		
QY 590	tgaaacttgaaagccttca tcga -cgacgcggtctgcgcgcggtctatactccggtgcgagtgc	648		
Db 1065	tcGAGCTGAAGAGACTCTGCTCGACCAAGGCCGCTGACACCAGCGGCTTGCCACCCGGAGCTGG	1124		
QY 649	tatctgcatacggttgagaaacgggtctcgaaactctggaaggcgggggtccggtctcgaaagc	708		
Db 1125	TACCTCACACACATTCACAGCGGGGCTTCAGACGCGTGGAGGGCGGACACCGGTCGTGGCCGTG	1184		
QY 709	gcgcattttcttgtaaggtgcga 731			
Db 1185	AACCTGTTCTCTCCGCGCGTGAA 1207			

[illegible]

QY	105	ggggcgctaacccggggtgatacaacaagatctggggcgccggaagccggccaggtgcatggagt	164
Db	747	GGGCGCGTATGTGGTCCAGAAACAACCGCTGGGGCCGGAGTGGCCCCCGAGTGGTACCGC	806
QY	165	cggactggaaaagggaacttcacgatcacacggcgccgatcacgaaacaacggt	224
Db	807	TACGCAAGCGGCTTCGGGTGACCCAGGCGGCGCGGTCCACGAACGGTGCC	866
QY	225	ggcgccatccggccatctactctgggtgtgccactggggcgccctgcagagcaatcggg	284
Db	867	GAAATCCATATCCCTCGGCTTCACAGGGGTGTCATCACCAATGCTGCGCCGGGACGAA	926
QY	285	atggcgcgcgcgcgagagagctgtccgacgctggcgacgagctgagagctcagcgcat	344
Db	927	CCTCCCGGCGCAGGTGACCGGTATGCCAGGCCCGCCACACATCTCTAGGCTTTGT	986
QY	345	cacgacggcgccgtggaatgcgccttaagacaatctgtgttcagttcccgctcaagaattccg	404
Db	987	CGGCAAGCCCGGTATCAACGGGTGTACAGATCTGGGTGGAGCCCGACACCAACGAAGAA	1046
QY	405	caagcgctacagcggcgcgccgagctgtatgtctgtcgtgaacttggaaacggcgcggtat	464
Db	1047	CGGGGTGAAC-----GTACCGAATCATGTATGCGGTCAACAAGTGTGGCGCGATCA	1100
QY	465	ggcgggcgccagcgcgcgtgycgcaacgttgaacttggcgccgagcttgggaattctgtg	524
Db	1101	GCCCATTCGGCTTCGACGGCGCGACCGCTCCGTGCGCGGGCGCACCTGCAGTGTGGAG	1160
QY	525	tgcgactggtgagcttgaattatcatgccttacccgctcacgacgcgcacacagctggag	584
Db	1161	GGGAGCAACGGCTCCACAGACGATCATCTCTTCCTTCGCGCCCTTGCGCGTGCACGCTG	1220
QY	585	cgacttgcgaacctggaaggccttcattcgaagcgggtgtgcggcgcgctatccgcgcga	644
Db	1221	GAGCTTCGACGTCATGAGCTTCGTCCGAACAACCATGCCCGCGCATGGCGCAACAA	1280
QY	645	gtgagtactgcacgtggagagcgggcttgaactctcggggagggcgcgctgtcg	704
Db	1281	CTGGTATCTCACACATGTCACAGCAGAGATTCAGCGCGTGGCAGAACGGTGGCGACTGC	1340
QY	705	aagcgcgatatttcctgtaacggtgca	731
Db	1341	GGTGAACCTCTTCCTCAACCGTGA	1367
RESULT	9		
SREGS			
LOCUS		1470 bp	DNA
DEFINITION	S.rochei egls gene.		linear
ACCESSION	X73953		BCT 16-DEC-1994
VERSION	X73953.1		
KEYWORDS	cellulase; egls gene; endoglucanase.		
SOURCE	Streptomyces rochei.		
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomycetaceae; Streptomyces.		
REFERENCE	1 (bases 1 to 1470)		
AUTHORS	Mastromei, G.		
JOURNAL	Direct Submission		
TITLE	Submitted (07-JUL-1993) G. Mastromei, Università di Firenze, Dipt di Biologia Animale e Genetica, Via Romana, 17/19, 50125 Firenze, ITALY		
REFERENCE	2 (bases 1 to 1470)		
AUTHORS	Petito, B., Hannhart, E., Irdani, T., Iqbal, M., McCarthy, A.J. and Mastromei, G.		
TITLE	Characterization and sequence analysis of a Streptomyces rochei A2		
JOURNAL	Journal of Molecular Biology		
MEDLINE	Gene 148 (1), 119-124 (1994)		
FEATURES	95011642		
FEATURES	Location/Qualifiers		
SOURCE	1. 1470		
	/organism="Streptomyces rochei"		

```

repeat_unit      /strain="A2"
                  /db_xref="taxon:1928"
                  /clone="E.coli pCSF1"
                  90. .103
repeat_unit      /rpt_type=INVERTED
                  112. .125
repeat_unit      /rpt_type=INVERTED
                  152. .157
RBS              166. .1314
gene             /gene="egIs"
                  166. .1314
CDS

```

stem_loop	1321	1340		
stem_loop	1351	1403		
stem_loop	1419	1453		
BASE COUNT	211	a	579	c
ORIGIN			475	g
				205
				t

Query Match	13.1%;	Score 96.2;	DB 1;	Length 1470;
Best Local Similarity	47.3%;	Pred. No. 7.2e-05;		
Matches 327;	Conservative 0;	Mismatches 358;	Indels 6;	Gaps 1;

QY	41	cgagagccggaagcccgagaccgagaccgctgctgctgagacgctggaagacgcgcgagctg	100
Db	254	CCGCGACCGCCCCCGCCACAGGCGCGACACAGCATGTGGGAAGATTTCCGGCTTCGACCGCTCA	313
QY	101	ccgagggcgctgctaacgggtgatacaaaagatagggcgcgagaaagacggcccgagtgat	160
Db	314	TTCAGGGGCGCTTTACCTCCTCCAGAACCAACCGCTGGGGACACAGCCCGACCCAGTGGCTCA	373
QY	161	aggtcgagcctggaagaaacgggcaacttcagatcatcacgcggcgcgatcatcagacaagggcaaca	220
Db	374	CCGCGACCGAGACTCCGGCTTCCCGGGTTCAGACAGCGCGAGCGGCTCGGTGCCGACCAAGCGCG	433
QY	221	aagttgacggcctatccgagccacatacttcacgtgtagcacttgaggcgcgcttcgacagaaat	280
Db	434	CGCGCAATCTGTACCCCGTGGGCTTCACACGGGCTGCATACACGAATGTTGCCGGGCA	493
QY	281	cggagattgcgcgcgcggtgctgcaagagcgctgctcgaagttgcgacagagctcggaagctcaagc	340
Db	494	CGGGCTCTCCCGCGGGGATCAAGCGGCATCTCTCCGGCCCGCACAGCAATCTGTAGCGCT	553
QY	341	cgatcacgacgggcccgcggaatgctgcgactacgacatctggtctcaagttccggtcacaagat	400
Db	554	TCTGTGACAAACGCCGTGTACAAACGCTCTGTAGACATCTGGCT-----GGACCCGAGCG	607
QY	401	ccgggaagaagctacagcggcgcgcgcccgagctcgatgatactgctgtgaactcggaacggcgcgcg	460
Db	608	CCCCGACCGACGCGGGGTGAACCGGACCGGAGATCATATCTGTCTTCAACCGGGGTGGGGCAGA	667
QY	461	tgatgtccggcgcgacgacggcgctggtccacagctgtaaatctggccggggccacacttgaaagttc	520
Db	668	TTCACGCCGATGCGGCTCGCAGGTGCGGCACAGCGCCTCCGTGGCGGGCGGCACACTTGGAGGTGT	727
QY	521	ggtatgcccgaacttgggaactcggaattatcatcgctctaacggcgcgcaagcccgccacacgctcgg	580
Db	728	GGTCTCGGGCGGAACCGCACCAACGACGATGTCTGTCTCTGCTGCTCTTCCTCCGCGGATGACGA	787

Qy	581	tgaacgagcttggaccttaaggccctcaatcgacagcgagctcccgcgagctaatccgcg	640
Db	788	gctgacagcttgcagctgattgatgaccttgcctccggcgacacctgcgcccgcgccctgcggcgca	847
Qy	641	cggagctggtatctcgtcgtctggagacgagcttcgaaacctggagagcgaggcgctc	700
Db	848	acgacttgctaccttaacgacgacatccagccggcttccgagcccttgacacgaacgctgcggctc	907
Qy	701	tgcgaagcgcccaattttccgtaacggtgca	731
Db	908	tccgggtcaactcctctctcgcgacccgtgaa	938

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
AF130408	1311 bp mRNA linear BCT 14-APR-1999	AF130408	AF130408	AF130408.1 GI:4583444	Streptomyces viridosporus.	Streptomyces viridosporus.	1 (bases 1 to 1311)	Ramachandran, S. and Crawford, D.L.	Characterization and sequence analysis of two genes involved in cellulase degradation in Streptomyces viridosporus T7A, and its expression in Escherichia coli	Submitted (24-FEB-1999) Department of Microbiology, Molecular Biology and Biochemistry, University of Idaho, LSN 125, Gibb Hall, Moscow, ID 83844-3052, USA

FEATURES	source	Location/Qualifiers
gene		1..1311 /organism="Streptomyces viridosporus" /strain="7A" /db_xref="taxon:67581" 1..1311 /gene="celS1"
repeat_region		82..97 /function="putative regulatory region" /rpt_type=inverted
RBS		144..149 /gene="celS1" /note="putative"
CDS		157..1311 /gene="celS1" /note="celS1; family H member" /codon_start=1 /transl_table=1 /product="cellulase" /protein_id="AAD25090.1" /db_xref="GI:4583445" /translation="MRRITAPDPAPCAVMSPPCPSPSSRRMPRRRRRAODTLCEP IGTTIGRIYVQNNRWSSSPQCVATDITGRLOADGSPVFNARPSISVNGCH YTNQSPGKLPARISGISAPARIXGISAPSSISYGVGAVYNASDIWLADLPRT DGNVREITMIFENKVGPIQPIGSOVGTAIVGAGPQMGAGLAVNSFSIVDVGCGT MSFPGVFEVRETVARGAQNNDMYLTSVQAGEPQMGAGLAVNSFSIVDVGCGT PGTPAACTVATNVMPPGFPANTYVGTGSSAVDGMGLAFTLPAGORTTANMAVYR PLIHGQAGSGPDNNTGSRRRKPDPRVPGNIGRHILRAERFQLNGTACTTA"
misc_feature		166..168 /gene="celS1" /note="rare tta codon" /function="putative regulatory function"
BASE COUNT	216 a 494 c 414 g 186 t	1 others
ORIGIN		

[illegible]

Query Match	Best Local Similarity	Matches	298; Conservative	12.0%; Score 87.6; DB 1; Length 1565;	47.5%; Fred. No. 0.0012;	0; Mismatches 324; Indels 6; Gaps 1;
QY 105	ggggcgctacccggtgatacaacaagatgagggcgcgagacgcgcagtgcatgagtc	164				
Db 489	ggggcagatcagctgctcagacacacacacgctggcgctcaccgccccacatgcgtcagggc	548				
QY 165	cggacttgaaacggggaacttcagatcacacagggcgcgatcaccgaacagcgacaacgct	224				
Db 549	caccgacacacggccttccgggtaccgacgacgacgctcggcagccagccagccacggggagccgc	608				
QY 225	gacgcgctaccccgcatctacttcgggtgacacatcgggagcgctgcacagagaaatctggg	284				
Db 609	gaagtcctgacccgctggcttcaacaggcgctccacatcaacagactgtttacacggcgacgga	668				
QY 285	attgcgcggcgcggtgcagagagctgltccgaagtgccgacagagctggaagctcagcgcat	344				
Db 669	ccctcccgctccgctcgcagacacccgctccggcgccgctccacacatctcgtacgcttgcgt	728				
QY 345	cacgacggcgcgctggaatgacgctcagacatctggttcaagttccggtcaagaaatccgg	404				
Db 729	cgacggcgccgctctacacacgcttctgtagcacatattag-----ctggacaccgacggccgc	782				
QY 405	caacgctacacagcggcgcgagctgtagtctggtctgaaatcgtgaacgycggcgtagat	464				
Db 783	caccgacggggtagaacacagaccgaatcatgtctggtcaacaggggtgggtccgatcca	842				
QY 465	gacggggcgagcggcggtggccacggctggaacttgccggggggccacatcgggaagtctgta	524				
Db 843	gcccatctggcctacacgggtggcagcggccttccgtccggcgcccgacatgggaagttggag	902				
QY 525	tgcgcgctgagactggaattatcatcgcttaccggcgacagacgacccacacagctcgtgtag	584				
Db 903	cggcggaacacggcctcgacacgacgctgctgcttgcggacacgctggcgatcagggcgctg	962				
QY 585	cgaagcttgagacctgaagcgcttcatcgaagcagcgggtcggcgcggtatcatcctggcgga	644				
Db 963	gaagcttgacgctgaagatttgcgtccggcgacacgctggcgacggaatccgcgaatacca	1022				
QY 645	gtggtatctgcacgtcggtgagaagagggctctgaaactctggtgagggcgggcgccggtctcg	704				
Db 1023	ctggatcctcgacgacgcttccagggcggttccagccctgacgaaacggcgcccggaactggc	1082				
QY 705	aagcgcgatcttccgtaacggtgcag	732				

Db	1083	CGTAGACCTCTTCCTCCACCGTCGAG	1110
RESULT	12		
SCG11A			
LOCUS		41782 bp	Linear
DEFINITION		DNA	BCT 21-MAR-2000
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
AUTHORS			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

SCG11A 41782 bp DNA Linear BCT 21-MAR-2000

Streptomyces coelicolor cosmid G11A.

AI133210

AI133210.1 GI:6468421

acyl-CoA dehydrogenase; aldehyde dehydrogenase; amidase; dihydroxy-acid dehydratase; DNA polymerase III beta chain; export protein; gntR-family transcriptional regulator; lacI-family transcriptional regulator; marR-family transcriptional regulator; oxido-reductase; reductase; regulatory protein; secreted cellulase B precursor; secreted cellulose binding protein; tetr-family transcriptional regulator; transcriptional regulator; xylose repressor; xylose kinase.

Streptomyces coelicolor A3(2).

Streptomyces coelicolor A3(2)

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

1 (bases 1 to 41782)

Redenbach, M., Krieser, H.M., Denapate, D., Elchert, A., Cullum, J., Kinashi, H., and Hopwood, D.A.

A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

Mol. Microbiol. 21 (1), 77-96 (1996)

97000351

2 (bases 1 to 41782)

Oliver, K. and Harris, D.

Unpublished

3 (bases 1 to 41782)

Cerdeno, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.

Submitted (23-NOV-1999) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.

(URL: [http://www.sanger.ac.uk/projects/S\\_coelicolor/](http://www.sanger.ac.uk/projects/S_coelicolor/))

CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database

The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nuclear Acids Research, 22(122):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nh.gov.jp/jun/cgi-bin/frameplot.pl>

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-11bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid G11A lies on the AseI-6 genomic restriction fragment.

Location/Qualifiers

1..41782

**gene**

/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid F37"

**misc\_feature**

1..1299  
/gene="xylB"  
1..101  
/gene="xylB"  
note="nominal overlap with Streptomyces coelicolor cosmid 2StG11"

**CDS**

<1..1299  
/gene="xylB"  
/note="SCG11A\_01, xylB, xylose kinase, len: >432 aa;  
highly similar to SW:xylB\_STRRU (EMBL:M75789) Streptomyces rubinosus xyloxyase kinase (EC 2.7.1.17) (xylosekinase)  
xylB, 461 aa; fasta scores: opt: 2553 z-score: 2795.9 E():  
0; 89.2% identity In 427 aa overlap. Contains match to Pfam entry PF00370 FGCGY, FGCG family of carbohydrate kinases and two Prosite matches to entries PS00933 FGCG family of carbohydrate kinases signature and PS00445 FGCG family of carbohydrate kinases signature 2"  
/codon\_start=1  
/transl\_table=11  
/product="xylose kinase"  
protein\_id="CAB61582.1"  
/db\_xref="GI:6468422"  
/translation="DPROMMDALGEALSGCEEARAAAVSVGGQHGIVTLDRGPSPVAPLAMDVSASPAQRILDELGAQMARTGSVPASTVTYKMMATLEHPEEARKAKARLRHDHYTERLTGEGTTDGDVGCTGMWAGCPVADELLAYDLRALPPLRVARGVEACTGVRDGHGLPRPSKGTLYLAAGTGDNAAAALIGLRPGYPNWSLGTSRTAYVSSRRPADTVAGGFADARGDWMLPACTLNCTLVADVNASLLGIDRAVEGDTITVLPLELDGERPNLPFHSSEGLHLGRHDTTQAQLQAAYDGAVSHLGALLDVLDADPPSAPLLDIGGARPTNGMOOTVRRLSGRPVOIPEARELIVALGAQAAGLTGEDNAVARRRNTTAGPVDAVERDEATTINRITGYLSDAPILERDAASR"  
76..1122  
/gene="xylB"  
/note="PFam match to entry PF00370 FGCGY, FGCG family of carbohydrate kinases, score 514.30, E-value 8.7e-151"  
250..288  
/gene="xylB"  
/note="PS00933 FGCG family of carbohydrate kinases signature 1"  
874..936  
/gene="xylB"  
/note="PS00445 FGCG family of carbohydrate kinases signature 2"  
1519..1523  
1525..2733  
/gene="SCG11A\_02"  
1525..2733  
/gene="SCG11A\_02"  
/note="SCG11A\_02, possible xylose repressor, len: 402 aa;  
similar to SW:xylR\_BACSU (EMBL:M7248) Bacillus subtilis xylose repressor xylR, 384 aa; fasta scores: opt: 612 z-score: 642.4 E(): 2.1e-28; 29.4% identity in 381 aa overlap. Contains match to Pfam entry PF00460 ROK, ROK family and a possible helix-turn-helix motif at residues 43..64 (+3.15 SD)"  
/codon\_start=1  
/transl\_table=11  
/product="putative xylose repressor"  
protein\_id="CAB61583.1"  
/db\_xref="GI:6468423"  
/translation="MSAPPHEAOAPRKPRALPDTOQGRRNRNSRWHTYSABEPSLRSAAVASRIETRAAVALSTLDELIRSLLEELDERPERGVRGPSALAASGOQPATIGEAEVGNLAWCAVDLRGRVARARYXSNRGSRSEPYEDQTGLVROVVSOATELEMPAGAIAVPCISVAIEDGRTVVARPLYDMNDADIALLPADIPRYDNEANFGALAEIMLEDGTPRFPELVNIIEIGAIVVEDGRLRCTRFPAGELEGHPVNPHGPCACAGRCRCLSEVGEKAIVRAAGVEPEGEDRGVGLGRAAEQGEDVRALREMGTAIGALTGANVLLDDPEVYLGALACLAGLPWLFLSLDELARRTAGCAPVAASELPQGPLGAAHSVRYAVLDDPGVAIVEKA"  
1822..2373  
/gene="SCG11A\_02"

**misc\_feature**

1822..2373

/note="Pfam match to entry PF01510 ROK. ROK family, score 131.80, E-value 4e-38"  
3041..3646  
/gene="SCG11A.03"  
3041..3646  
/gene="SCG11A.03"  
/note="SCG11A.03", probable amidase, len: 201 aa; similar to SW:AMPD\_ECOLI (EMBL:X15237) Escherichia coli anhydro-N-acetylmuramyl-tripeptide amidase AMPD, 183aa; fasta scores: opt: 211 z-score: 267.1 E(): 1.7e-07; 39.5% identity in 114 aa overlap and to Streptomyces coelicolor SC2A11.21c; fasta scores: opt: 674 z-score: 751.6 E(): 0; 50.0% identity in 202 aa overlap. Contains match to Pfam entry PF01510 Amidase\_2, N-acetylmuramyl-L-alanine amidase"  
/codon\_start=1  
/transl\_table=11  
/product="putative amidase"  
/protein\_id="CAB61584.1"  
/db\_xref="GI:6468424"  
/translation="MERKRTAPVPTRRRLKGAATATVPYTLTSGRAAQQRAVDYP  
SAEWLPASTSNYSRSPRPAYPVDFVYVHTQETVADTSLFRDPKQVSAHYVRRS  
DGHVACVRETDIAHAGNMWDYNTSGICEHGGWDRPFPNMAEGRALTAACIT  
AYGTPKDRPHITAHHEVPEGSDHTDPGPFMDWTRYIRLVNFA"  
3191..3637  
/gene="SCG11A.03"  
/note="Pfam match to entry PF01510 Amidase\_2, N-acetylmuramyl-L-alanine amidase, score 89.80, E-value 5.7e-23"  
3811..3815  
3821..5080  
3821..5080  
/gene="SCG11A.04"  
3821..5080  
/gene="SCG11A.04"  
/note="SCG11A.04", possible transcriptional regulator, len: aa: similar to TR:O31551 (EMBL:299108) Bacillus subtilis transcriptional regulator ACOR, 605 aa; fasta scores: opt: 353 z-score: 446.5 E(): 1.7e-17; 34.7% identity in 199 aa overlap"  
/codon\_start=1  
/transl\_table=11  
/product="putative transcriptional regulator"  
/protein\_id="CAB61585.1"  
/db\_xref="GI:6468425"  
/translation="MTDAWLALPEADPYERARALRAHETFEAGTVRRPVRAVAD  
SWRSVRAGVPGDGTASVELMDGLGAYRAEPLRVMPLVELLGTFAADEHLAY  
CDAGRLLMVEGHPATRRRAGRMNFPGARMSGAVGTAPAAVAGRPVFAAEH  
FIRVQPMTCAPAVHDPRGTAVLGAVIDTGGDLAHPHSLGFVOAVARAASOLAL  
TPERSASEAETALGRBALISADGRARVRLRRHSEIVLVLAHHPEGLTGELICAL  
VEDDEVPTVTLAEALRLRGITIGRGLASRPRLMPYSPVVERRLRAAYRTAA  
AAAYAPLPGSQAPVAGRLRRRLADGLRALIALTCGDPDLADMAHAPMGEDLDVWRA  
LAAVPTATSTSRLLAESELAAPNW"  
5205..5209  
5218..6741  
5218..6741  
/gene="thca"  
5218..6741  
/gene="thca"  
/note="SCG11A.05, thca, aldehyde dehydrogenase, len:534 aa; highly similar to SW:THCA\_RHOSN (EMBL:U17129)  
Query Match 12.0%: Score 87.6; DB 1: Length 41782;  
Best Local Similarity 47.5%: Pred No. 0 00047;  
Matches 298; Conservative 0; Mismatches 324; Indels 6; Gaps 1;

Db 27948 GAAGTCGTAACCCGCTGCTTTCACAGCGCTCCACTACAGCACTGTTACACCGGACGCGC 28007  
Qy 285 atggccgagcgcgctgcagagagctgtccagacgctgcgacagagctgtgacgtacgcgcgat 344  
Db 28008 CCTCCCGCTCCGCTCGACACCGCTCTCCGGGCGCGCTCCAGCATCTCTGACCGCTTCGT 28067  
Qy 345 cagcagcgcgctgtgaatcgcgcacagacacatcgtgttcagcccgctcacatccg 404  
Db 28068 CGACGGCGCGCTTACACACCGCTCTGACGACATATGG-----CTGAGCCGACGGCGCG 28121  
Qy 405 caacggtctacagcgcgcgcgcgagctgacgtgacgtgacgtgacgtgacgtgacgtgac 464  
Db 28122 CACCGACGGGGGTGAACACAGCAGCATGATGATGATGATGATGATGATGATGATGATGAT 28181  
Qy 465 ggcggcgagcgagcgcgctgtgcacacgtgtgaacgtgcgcgcgcgcgcgcgcgcgcgcg 524  
Db 28182 GCCATCGGCTCACCGGTGGGACGGGCTTCCTCGGCGGCGGACCTGGAGGTGTGGAG 28241  
Qy 525 tgcgcactggaactggaatcacatcgcctaccccgcgacagcgcccaacacgcgcgtcgag 584  
Db 28242 CGCGCGCAACGGCTCGACAGCAGCTGCTGCTCGTGCGGACCGCTCGGCGATCAGCGGCTG 28301  
Qy 585 cgagctggaactggaagcctcatcgaacgagcgcgctgcgcgcgctgacatccgcgcgga 644  
Db 28302 GAGCTTCGACGCTATGACTTCCTCGCGGCGGACGCTCGCGCGGACCTCGCGCGGACGACGA 28361  
Qy 645 gtggtatctcagctgagcggtgagacgagcggtcgaacactgagcgagcgcgcgctgctcg 704  
Db 28362 CTGCTACTGACGACGCTTCAGCGCGGGGCTTCGAGCCGTGGCAGAACGCGCGCGACTGGC 28421  
Qy 705 aagcgcgactttccgtaacggtcgag 732  
Db 28422 CGTGACCTCCTTCTCTCCACCGTCGAG 28449  
RESULT 13  
LOCUS AE005123 11202 bp DNA linear BCT 12-FEB-2001  
DEFINITION Halobacterium sp. NRC-1 section 154 of 170 of the complete genome.  
ACCESSION AE005123 AE004437  
VERSION AE005123.1 GI:10581826  
KEYWORDS  
SOURCE Halobacterium sp. NRC-1.  
ORGANISM Halobacterium sp. NRC-1  
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
Halobacteriaceae; Halobacterium.  
REFERENCE  
1 (bases 1 to 11202)  
Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
Shukla H.D., Lasky S.R., Baliga N., Thorsson V., Sirogna J.,  
Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
Leithausner B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
Maddocks D.G., Jablonski P.E., Krebs M.P., Angelino C.M., Dale H.,  
Isenbarger T.A., Peck R.F., Pohlshrod M., Spudich J.L.,  
Jung K.-H., Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P.,  
Omer A.D., Ehardt H., Lowe T.M., Liang P., Riley M., Hood L. and  
Dassarma S.  
From the cover: genome sequence of halobacterium species NRC-1  
Proc. Natl. Acad. Sci. USA 97 (22), 12176-12181 (2000)  
11016950  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Direct Submission  
Submitted (14-JUL-2000) Institute for Systems Biology, 4225  
Roosevelt Way NE, Seattle, WA 98105, USA



FEATURES		Location/Qualifiers
source		1..11202 /organism="Halobacterium sp. NRC-1" /strain="NRC-1" /db_xref="taxon:64091"
gene		76..1281 /gene="meta" /note="VNG2420c"
CDS		76..1281 /gene="meta" /note="Meta" /codon_start=1 /transl_table=11 /product="homoserine O-acetyltransferase" /protein_id="AAG20507.1" /db_xref="GI:10581827"
gene		/translation="MAALAVIAVHNITVAKNYOYCCTDGGPMPTTELDSGFPGCDTV DSLTVYATRYGDTGSNAVLVGHALTGSNHYGRADGDDRCWMPPIVSGFRPIDTTEY FYVCANVPSCGHTGTPPADGPGDPGWTAFPPVTVDPWTRCQRRLDTLGVRLHAY VGGSVGNMADLMARRYDDVHRIAIVATAPRLDQCLGLNARRAIKGDPMNDGD YVGAHPDRGLALRQLGHVYLSKSNADKFGRTATDTALPADPAEFPAEDVE SYLDQADTFVARFDANSYLLYLRAMDADLAGRPVADALAFDGDALVLSFTGDM HTAQSATVADAFRAGGAAGVAHHVVDSDYGHDAFLVDSAAVGPRLADFLAAGTDGA ATTDAATGCV"
gene		1351..2646 /gene="hal" /note="VNG2421c"
CDS		1351..2646 /gene="hal" /note="hal" /codon_start=1 /transl_table=11 /product="O-acetyl homoserine" /protein_id="AAG20508.1" /db_xref="GI:10581828"
gene		/translation="MGHDGLHTNSVHAGORVATPGARAPLYOTTSYAFPSADAA GCPALERGCYITSRIMNPVETLQRLALBEGCVAVATASGMALDLATFLIARAGD SVVAASDLTGGTGYTLTHSAQRGVDTPLVDLDATADADADADATVAVLEVGVPS LTPDLERLADIANDNGVPLDVNTFATPALATPIDHGADIVHSTVYIHAAGTVG GALVADGSEFMDAHADYPEIQAENPAGVFTFRFGDAFTYAAIAAGLDLQOQ SPEDAMOTLOKLETLPLRMOGCRNNAQVAEHLRDPNVSWNYPLADHDHDAAT YLDGCVGGMILTFGVGDGYEAAQSVTEETLISLANVGDAKTLVHPASTHQQLTPE AORAGCVREPMRVSGIEDPADIVADELTAEAVGSA"
gene		2715..5753 /gene="g1cd" /note="VNG2422c"
CDS		2715..5753 /gene="g1cd" /note="g1cd" /codon_start=1 /transl_table=11 /product="glycolate oxidase subunit" /protein_id="AAG20509.1" /db_xref="GI:10581829"
gene		/translation="MATRDLRSDPSTPDADYVTSDSVERGLVDDLRERVDGR FSTYRQLYATDASTYELPVGVPFTSTDDVAAYVSTCADRGIPVLPFGGGSTLAGQ AVNEAVVLDLFRHAMNGFEGDEAEFAEQPCTTLAALNDELAAHLKAPAPPMGDK SVLGAIGNNSTKANGSLQYEKADGYLESAEVLADGTVTTLGWDVTDLDALADPGD LEARIYADLRLLDEADRIIRRFEPDLKRNVSQYINRDLDAEAHGAIVANARLWLAGS EETIGITAEVSLIEPVPAETAVAMLTVDVDAQODVAPVLHDPSANVAMWDVFLD LARDFDEPADYVGLPDPGDTSLVEFYADSPAERARRADILDRBPATTPAADPPG AATTPDAVHADLEAYDDROAHFWMKMKACLPILSTSDKEHPIYIEDPAIRPDH LPAFVADFOAILDDLDTVASYYAHAGPGLVHLPLVTLADEGVAKMAIADRIVDLH VYGGVAVSGEHDGARSQSNKRFGEDLMASFRTKAFDPMILNPFGVGDAAE IDMTDLHRYAPDDEPDGAPALNMDTENGFGMTLELCCAGCTGHOOTGCGMCP VYAAENETLSHGRANMLBSAMNGHLPADLPDEPLEYLDLCVCGKCKLDCPSGD MAKLTVEYKHAHQDGVGLDRLEGHTALPHLSALAPVSNMATALDGAALAEKT LGLAERDLPTFRRESLRAMFDRAGSRVPEAAETVLLADPYTNTSTVAVGAAY RVLEAGVVALAPDVGTSGRPAYSRGLIDTRRIARRENVAAALAPRVADGMPVYVEP SDAVVODDYHDLLSGAPAAVAAANTYGVCEYLDRIRGLDAAGFDAGDSESLVHGICH OKAQRDHHAGVLRAGYEVDLPDSTCCGMAGSFYEAEHY SMSQAINDILLADQVAD SDADVTYAGSSCSRQLSERDDVDTDGPAPPHVOKLADAL"
gene		complement(5763..6470)
gene		/gene="serB" /note="VNG2423c" complement(5763..6470) /gene="serB" /note="serB" /codon_start=1 /transl_table=11 /product="phosphoserine phosphatase" /protein_id="AAG20510.1" /db_xref="GI:10581830"
CDS		/translation="MLCWSAPKYCRMKCYVAGRRATPGMTLVAFPDGTLAESEMLDR IARAGGDEVAATITERMARKEBLSADSLREARQOIVAGIPESAAAYAVDGYRLRAGC DLVAKLRDGGVAVVLTGTFKPGVAAPDAAGVADVGVGNLVLADELGAEGPL VEGTDKDALRDACEAAGTTPAAAVAVGGADVPMIDAAAGTAIGVDPKPGVADHCDHT VSSMDALGRVLDDHGA"
gene		6555..8138 /gene="serA1" /note="VNG2424c"
CDS		6555..8138 /gene="serA1" /note="serA1" /codon_start=1 /transl_table=11 /product="phosphoglycerate dehydrogenase" /protein_id="AAG20511.1" /db_xref="GI:10581831"
gene		/translation="MRVLVTDPIADAGLRLRDAGHEVYTTAYDATGALDLDAVSDAHA LIVRSQTAVTDVAFEAAPDLIVARAGIGVNDIDDAATDHGVANANPAGVNRAAE HTVALAPAAARSIPOAHARLDAGSNAKDDYLGTELSGRTIGVYGLRGQEVATRLDS LGMDLVADDPYIGEDRANOIGAEVDIEVCVARADPLTIHVPLEPTEGGLIGEARLAR MDGCVIYVARGCYVDEDLADPADQVYIAGALADYFRTPEPLPAASPLHDASTITTP HIGASTKAQENAVATDQVVAALAGDPVVALAPSAERSAFDRIFFVDLAEATAG TYAAALFDRIERVDVITYQGEVADBDVALYTAAGQGAFAGLEQVNVNAPRVAER GJAVTEGHNHRESEAYOSLVSIVYVNGDAELTVSGTLFPGDDPRLVEIDGFVYEAAPG HMLVARNHPDTPGVIGFTIGVLTAVANIAGMNFARERAGGALTYYNLDADVPARALD ELAGDRLIVDVYSIELNCE"
gene		8200..8643 /gene="act" /note="VNG2426c"
CDS		8200..8643 /gene="act" /note="act" /codon_start=1 /transl_table=11 /product="acyl-CoA thioester hydrolase" /protein_id="AAG20512.1" /db_xref="GI:10581832"
gene		/translation="MGSTSDTFMENRERYOPDTNNVYASAHGCVVYKMWDEIGAMAAM RKAQTCVTARINSIDFERPVQGDICVIOATAYDTGHTSIVYRLARERDRSGETE PTTDSIFVFAVADDMDQPTAVPDLVGSERKELDEALAGHDE"
gene		8825..9886 /gene="dld" /note="VNG2429c"
CDS		8825..9886 /gene="dld" /note="dld" /codon_start=1 /transl_table=11 /product="D-lactate dehydrogenase" /protein_id="AAG20513.1" /db_xref="GI:10581833"
gene		/translation="MISALTADIHITNTKAGFAFTTLMGLVALLQYPSGRSLDALSRLK TLIVVGLAASGFFVVALNARSYAALLVGAALVGTGAGLYPTPAARALVADLYSTQGR AGFLHTASGDLGMSAAGIAVALALISRYAFVPIYCLTAAVAILALVWSESYVE RSLSLVDRDAAALLSGTOBRMLIVAYTYAFWQWIAFLPFTFDGKEFAFVYGTVA RGALFAVGAIAAPTKQTLSDTISRRLAVGALVIGTGAALSAVYVPSVPAANGAVYVF AAGLAFAPPMOSTLMDAFPPESAGDLGAMTIVYIGIGALGPTTYGVAVSTANYEIA FWGLGALLVAIVIIARVT"
gene		complement(9890..11137) /gene="thrc1" /note="VNG2430c" complement(9890..11137) /gene="thrc1"

```
/note="Thrc1"
/codon_start=1
/transl_table=11
/product="threonine synthase"
/protein_id="AAG20514.1"
/db_xref="GI:10581834"
/translation="MTDLSIAGETPACATDGVWLACVECGTIAFPADVVRKPECSG
LLEARYADPGFEAFSGRVRYADALPDVACSVQEGDTPLEYEAPDVAAAGVADV
VHEGMNPTGSKDRGMTGVAVAEELGDRILACSTGMTSALAAVGGRAFPVVLV
LPAGVAGKIAQALHGAIRICEVNDPDCDVAIELARCEAYLINSINPLEGO
KTIAREILEOSTRTATGEPRDRLVLPVGNAGNTALYKARELYAAGMTDDMPALIG
VDAAGAPVVEAVEAGRGTERMTWETRTAIRIGNPVNAKALPGVETGTAIVAV
SDRITDQRMALDGIIGVERPASASVAGLLALRREGISADERVVCILTTGHLKDP
AAAAAGCTTTPPADADGVLDAL"
BASE COUNT      1670 a      4414 c      3626 g      1492 t
ORIGIN
Query Match      9.4%; Score 69; DB 1; Length 11202;
Best Local Similarity 43.8%; Pred. No. 0.3;
Matches 300; Conservative 0; Mismatches 385; Indels 0; Gaps 0;
QY 18 cgaatgcgaacaagaaagaaacgagacgtgcccgaagccgaacgtgagctgtgcg 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7058 CGACCCCTCATCTGGGAGGACCGGCGGACAGCTCGCGCGAGCTGTCAGATCGA 7117
QY 78 acgctggagacgcgcgagatgtgcccgggggagctacacgggtgatcaacaacg 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7118 GACCTCGCTGCGCGCGCGGATTTCTCATCAGATCCACGTCCCATCAGCAGAC 7177
QY 138 cgcgcgaagaccgcgcagatgtgagctggaatggaacgggcaacttccagatca 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7178 CGGCTCATCTGGGAGGACCGGACACTCCCGGATGAGACGGGCGCTACGTCAAC 7237
QY 198 ggcgcgacacgaacaagctggccgcctacccgcaactactactctggtggcgca 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7238 CCGCGCGCGCGCTGCTGACAGAGACCGCTGCGCGACCGCCGACAGAGCGCT 7297
QY 258 ctggggcgccctgcagacgaatctgggagatgtccgcgcgcgtgcagagctgtc 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7298 GGGGGCGCGCTTCGACGTGTTCGACGAGCGCGCTCCGCGCGCGCTCCGCGTCA 7357
QY 318 ggcgaagaaactggaacgtcaacgcgaatcgaagcgcgctggaatgtccgctca 377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7358 CGCGACACATCATCATCACACGCGGCGCTCCACAGACCGCGCCGACGAGAA 7417
QY 378 ctggttcagtcctcgtcaagaattccgcaacgctcaacgagcgccgagctgatat 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7418 CGTCCGACACGACACCGCGGACCAAGTCTGCTCGCGCGCTCCGCGGACCGCT 7477
QY 438 ctgagctgaactggaacgcgcgtgtagtccgggagcgagcgctggtccagcttga 497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7478 CGGCGTGAACCGCGCCCGACGCGCGCTGCGCGCTTCGACCGCATCCGCCCTT 7537
QY 498 ggcgcgggagcaacttgggaagctgtgtatgctgcgaacttggaattacatgc 557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7538 CCTCGCGGAGACCGCGGACCGTCCCGCGCGCTTCGACCGCCCGCATTCAGG 7597
QY 558 ggcgaagagcccaacacgctcggtgagcagctggaacgtcctcaacgaacgc 617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7598 CGACGTCACAGTACCAAGGGAGAGGTCCGACAGAGACGTCCGCTCATCGCGG 7657
QY 618 ggtgcgcgcgcgtacatccgcgcgagatgtgatactgcatacggtggaagcgctcga 677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7658 CCAACAGGCGCGCTTTCGCGCGCTGAGTGGAGTCAACGCGGTCAACGCCCG 7717
QY 678 actcttggaaggcggtggtgcggtctgt 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7718 GGCGGAGGAGCGCGCATCGCCGCTG 7742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14  
AF335723

```
LOCUS      AF335723      4262 bp      DNA      linear      BCT 11-FEB-2001
DEFINITION Burkholderia pseudomallei formate hydrogenlyase subunit 4 and
ACCESSION  AF335723
VERSION    AF335723.1  GI:12744948
KEYWORDS   SOURCE
           Burkholderia pseudomallei.
           Burkholderia pseudomallei.
           Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
           Burkholderia; pseudomallei group.
REFERENCE  1 (bases 1 to 4262)
AUTHORS   Steiner,B., Meyer,R., Bowen,M. and Morrill,W.
TITLE     Random sequencing of Burkholderia pseudomallei strain G9313 for
          clinical PCR development
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 4262)
AUTHORS   Steiner,B., Meyer,R., Bowen,M. and Morrill,W.
TITLE     Direct Submission
JOURNAL   Submitted (12-JAN-2001) RRAT, Centers for Disease Control and
          Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA
FEATURES   Location/Qualifiers
           1..4262
            CDS
            source          /organism="Burkholderia pseudomallei"
                             /strain="G9313"
                             /db_xref="taxon:28450"
                             complement(89..1039)
                             /codon_start=1
                             /transl_table=11
                             /product="formate hydrogenlyase subunit 4"
                             /protein_id="AAK06856.1"
                             /db_xref="GI:12744950"
                             /translation="MVASGILSQTEILVALAAPLLTGWYNOGRAMLQNRRAISW
OYRMILHKLFNKESVAVAGAREPLERGAIVVWMAALCAIPIITSTELSPADAI
ALVGFALARVLSLAAMDIGAFGLTARREMLFLAEPLALVLSASLITSTL
LTSIVAAIGHRELATYPSLAEGIAFTVLSAENRCLPVNDPATILEMTHEALILE
VSGRIALMEMASIKIRAYSCIGIALMPGCIENARSGPAILALLALPALFKLLYGA
ALAVETTNARKRLERVPEFLASATFLAVIGLVHFLUGA"
            complement(1033..3039)
            /codon_start=1
            /transl_table=11
            /product="hydrogenase-4 component B"
            /protein_id="AAK06855.1"
            /db_xref="GI:12744949"
            /translation="MEGLLKLVCVFLAVALMLALCAGLALNRRTGYAHGLFPLGAL
AGLVLAALGAGLVAFGPSAVYPLGLPLDPLRPLADPLSAFLVGLVATGAVGAFSS
GYFRKGEGLAPGLICFEYHVCVAVSLVLDVADVCYFVMAETLTLTSAFTVMTNRI
AEIRRAAFYFLISHVGLALILFCGLAQAGGDTYFANMRIOHIDGLAASAEVLA
VGFAGKAGFPLHVMPLPAHNPAPSPVSAISGFVLKGIYGLITRTVDLHVQAM
GYMLALGLFGLALGAVSTIOTDMKRLLAYSSIDNGLMCVSLGLITLITFCYRLPSL
MALSTALYIOTVAHACFRSLFLCTGAVLHATGERNRGLRGLIRPMPVTRAMLVG
VAASAGLPPSSGEVSEMLLVDSLEFLETPGLPDSVLMTVPLVAAVLAALPAALYAMV
FEGIVFLGPRKALERADSPMERIAFGVMAASVLLGLAPVOYLVVLAALTRALV
GAGIATGNHMLLPAPVSTARASVYPALEMLFEVACALAMLLVRFYHGRIRAPV
SCGNPFTVARMODTAEFGOPRIREITFEFFERLIDPSFDROPAYRMSVADRTWHLL
YPTIALNRAVAIYGRLOTGRIATLYLTYSFVIALLTIVRRW"
BASE COUNT      744 a      1502 c      1477 g      538 t
ORIGIN
Query Match      9.4%; Score 68.8; DB 1; Length 4262;
Best Local Similarity 47.3%; Pred. No. 0.41;
Matches 208; Conservative 0; Mismatches 232; Indels 0; Gaps 0;
QY 41 cgaagcttagcccgacgacccgtgacgtgtgaggaacgtgcgcgcgagatgtg 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1404 CGCAGCAGCGGAGAGAGAACGATGAAGCGCGGCGATGTAGCTCGCGCGCGGTGG 1463
QY 101 ccggggggcgctaccgcggtgatcaacaacgtatggggcgcggaagccgcagtgcatg 160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1464 ACACGGGCGCGGAACACACACCGCTGATTGCTCCCGCATGCCCGCGGACAGCG 1523
QY 161 agtgcgacttgaaacggcgaattcagatcacagggcgatcacgacaacgggaaca 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 1524 CGCGCGTACGATGACGACCTTACGACTGACGAGCGGCGGAGCGGCGGACGACGCG 1583  
OY 221 acgtggccgcctatccggccatctactcgggtgctgacactgagcgccctgacagcaatt 280  
Db 1584 TCGCGGCGCGGACGACGACGACGATTCGCGTCCGCGGCTCGCGTCCGCGGCGGCT 1643  
OY 281 cgggattgcccggcgcgctgagagactgtcgaagctgcaagctgcaagctgcaagctcagc 340  
Db 1644 CGAGCTTGGCTCGCGGCGGCGGCGGAGGACGAGAGCGGAGGACCTGACCATCGCGT 1703  
OY 341 cgcatacgaagggccggctggaatgcccgcctacgacatctgttcacgtcccgctacgaatt 400  
Db 1704 AGCCGCGGACGCGCGCTGTACGCGGACGAGCGCGCGGACGAGCGGACCGCTCATCCCGA 1763  
OY 401 ccggcaacggctacagcgcgcgccgagctgatactgtcgtgaactggaagcgcgcg 460  
Db 1764 GCACCGAATCGGCGCGCGCGGTGANAAGAAAGCTCTGACAGAGACGATTCGAGACA 1823  
OY 461 tgatgcggcgcgcgcgcg 480  
Db 1824 CGAAGCGCTCGAGCGCGCG 1843

## RESULT 15

AE006992

LOCUS AE006992 14483 bp DNA linear BCT 27-Apr-2001

DEFINITION Mycobacterium tuberculosis CDC1551, section 78 of 280 of the

ACCESSION AE006992 AE000516

VERSION AE006992.1 GI:13880691

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

SETFLIATROEVSOGKLEVYSPNSPLGALIDAKVGETRSTYTPNGSTVSTVLSAE  
PYHS"  
complement(746. .1207)  
/gene="MT1112"  
/complement(746. .1207)  
/gene="MT1112"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK45368.1"  
/db\_xref="GI:13880693"  
/translation="MHPAGVNPQMTPTTPRPDARYGRPLSRARRRVAALGVVA  
AAGIVAVIGYORISTSAVTSIGYRLVDEBTASVTISVRSDESRPACTVRAT  
NSENCRRLVLPPESEATTVQYTTVKSQPMADVCGTEVPSYLRP"  
1282. .2148  
/gene="MT1113"  
1282. .2148  
/gene="MT1113"  
/note="similar to PID:699274; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ImbE protein"  
/protein\_id="AAK45369.1"  
/db\_xref="GI:13880694"  
/translation="MSELRLMAVHAHPDDESSGAATLARYADEGHVLYVLTGGER  
GELINPAMDLPVHGRIAEIRDEMTKAEITGVETWLGYPKDLPPPLPD  
CFARVPEVTEALVVRERPHVVTYTDENGGVPHPHICHOVVAAYAAADGC  
RPDAGEPWTYKLYVHGFLERPMQIDEPARHQRGPRPOMLAXHPDPLFSR  
VTRVDCSTYFSQRDALARAHATQIDPNAEFFRAPLAWOERLPTFEELASRLPAR  
PEPTELAGIEP"  
2145. .2411  
/gene="MT1114"  
2145. .2411  
/gene="MT1114"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK45370.1"  
/db\_xref="GI:13880695"  
/translation="MNQILSYIAEGCPGNTGPRDKASPVGLIVLVLTPLVR  
SMNQILKVPKSRDHPDLDADEGTDGPARPPPHBSG"  
2398. .4419  
/gene="MT1115"  
2398. .4419  
/gene="MT1115"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK45371.1"  
/db\_xref="GI:13880696"  
/translation="MSPANSGTNTLATSPYLROHADNPVHWOQWTFOLAEAAAR  
AVPILSVGACCHVMAHBSFDDDEVAAMNAFCVICKVDERPQIDVYVNAAT  
VALTGGGMPKTCFLTPNGRPFCGTYKKAFLDLSAISTWERRAEVQASDHI  
AAELRSMASGIPGGPEVAPELDCDDVAVALRODPAHGFAGAPKPEFSALELMR  
HYERTSPALFAVARTGNAMARCGIYDLDGGFARVSDGAVVPHPEKMKYDNLAL  
LRAVHARRTGDPRLARVAAOTARFLDELGSKAPADMTFSSLDADRGRCSTYV  
TPVQLTEVLGGDGRMAAEVFGTEAGTGHGTSVQLPADPDDAARDLVRALLVA  
RLARQPARDDKVTYSWNGLATLALAEASVALDDPALAHARCATRLDLHVVGRRL  
RASLGGVYDSALIEDHAMLATGLALYQLTSBEGALTAATGLDTPVAHFVGPOR  
PGRWPTADDAERLMRPSDPLDGATPSSASIAEALITAGHVVGBARERVWOLAAD  
TLRAHVILARAPRSAGHVLVAEAVRSGASSIAEALITAGHVVGBARERVWOLAAD  
VSGAAGSSALLVGRDVRVAGADAAYVRCGRVCDLPTVSAELATFALGVC"  
4429. .4548  
/gene="MT1116"  
4429. .4548  
/gene="MT1116"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11

gene  
CDS  
/transl\_table=11  
/product="transcription elongation factor GreA"  
/protein\_id="AAK45367.1"  
/db\_xref="GI:13880692"  
/translation="MTPDYTWLTQSHRRLKAEELDOLIANRPVIAEINDREEDL  
RENGTHAAREEGQOEARILODLILSNKVGAPKQSGVALPGSVVYVYNGDKSD

gene  
CDS

/product-"hypothetical protein"  
/protein\_id-"AAK45372.1"  
/db\_xref="GI:13880697"  
/translation="MAPVREIRDAVSACRVAIFTLATALGVPQONTNHIA"  
complement(4518..5246)  
/gene="MT1117"  
complement(4518..5246)  
/gene="MT1117"  
note="similar to PID:887849 GB:U00096 PID:1789266;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product-"Hly-III family protein"  
/protein\_id-"AAK45373.1"  
/db\_xref="GI:13880698"  
/translation="MSGADRTTTEAETPAHAHHLYEGVARVLTPERRGWTHYS  
ACTAVLAGASLVAVSMAVGSAKAGLTTLAIVAAITMETVSATYHRVNWSATARMMW  
KADHSMEFVEFTAGSVTFPALALPABHGRLVLSVMOGAIAGILMKCPAPRSYG  
VLVLLIGLGVAVWTATILLHNAGVALVLEFGALVSIGIILYAVRMPPWPPTFGY  
HEEFNACTVAIVACHIMMFVF"  
5357..6145  
/gene="MT1118"  
5357..6145  
/gene="MT1118"  
5357..6145  
/note="similar to SP:p20182 PID:1196916; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product-"undecaprenyl diphosphate synthase"  
/protein\_id-"AAK45374.1"  
/db\_xref="GI:13880699"  
/translation="WEIIIPRLKEPLRYLYELRLRGILAASKDLPRIHVIYLCGNRR  
MARSQGDVSYGYRPMGAAKIAEMLRMCHEGEIELATVYLSTENLGRDEPLALEE  
ITIDVVEECIAANMSVTRVGGDLGLEIEEPARLRGVESTPEVASFHVAAYGGG  
REIIDVAVALSKELANGATAEEVDVAETVEGISENLVTSGDDPDVIRTSEQL  
SEFLIMOSVSEMTEEAHPAFRHVDLRALRDYNSAHRSYGR"  
6283..8591  
/gene="MT1118.1"  
/note="similar to GB:ALJ23456; identified by sequence  
similarity; putative"  
8768..9088  
/gene="MT1118.2"  
8768..9088  
/gene="MT1118.2"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product-"hypothetical protein"  
/protein\_id-"AAK45375.1"  
/db\_xref="GI:13880700"  
/translation="MPCVGYGDRREFVNAVAVEAICENTLSCGPDPDVLYRTSGEOR  
LSGHGPGGGVSRRLRALRDYSTPHASTIPYPPIKSDGIASHRLAVSFEDALAG  
RVEL"  
9223..9675  
/gene="MT1119"  
9223..9675  
/gene="MT1119"  
9223..9675  
/note="similar to GB:ALJ23456; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product-"pe family protein"  
/protein\_id-"AAK45376.1"  
/db\_xref="GI:13880701"  
/translation="WGKLMSYMIATPPALTTAATDIDIGISAVSANAAYAATTG  
VLAAGDELVAIALRFNANAEEYHALSAQYAATQTLFVRITNGCGCFRRRKROCV  
TAAEHRAAGAGRRQRRRSSDGOWRLRQQRRHFCCGGCPFEFRHSEHRR"  
9692..9859  
/gene="MT1120"  
9692..9859  
/gene="MT1120"  
9692..9859  
/note="similar to GB:ALJ23456; identified by sequence  
similarity; putative"

[illegible]

Search completed: August 19, 2002, 12:45:06  
Job time: 10450 sec

```

/codon_start=1
/translation="PE family protein"
/protein_id="AAK45377.1"
/db_xref="GI:13880702"
/translation="MTTASATASSGVGDGIAATYAVASQWDGXYVANNTITGRDGF
Query Match      8.6%; Score 63.2; DB 1; Length 14483;
Best Local Similarity 46.8%; Pred. No. 1.8;
Matches 304; Conservative 0; Mismatches 338; Indels 8; Gaps 3;
QY      81  ccggagacgcgcgcgagatgtygccgggggggcgcaccgcgggtgatccaacacgatagggc-g 139
Db 10132  ctAGTTCCGGGAGAGACTACGCGTACCGGGAGAGCGTACGCGTGGCGAGAACACGCGTGGACACA 10191
QY      140  cggagacgcgcgcgcagtcgacttgagtcgcgaactggaacagggcgaattcaacatcacacaggg 199
Db 10192  CCCCCGCCGGGAGACCGTCACACGTCAGCGCAACCGGGTTCACCTTCACCGAGATGAACG 10251
QY      200  ccgataccagacagcagcgaacacagctgycgcgcctacgcgcacatactactcgggtggtccact 259
Db 10252  GTCGTCTCCCAACCAACGCGCCGCCGCTGATGATTCACCTCTGATCTTCCGAGGCGCTGCACT 10311
QY      260  ggagg-g-cgcctgcacgagcaaatcgcggatgctccgcgcgcgcgtgcagagctgctcgcagctg 318
Db 10312  GGGGGCCACCTCTCGTGGTGAGCACCAACCTGCTACTGAGTGGGCGCAATTTTGTGGCGG 10371
QY      319  cgcacagagctgagcgcctacgcgcgaatacagacagggccgctggatgctgcgcctacgacac 378
Db 10372  CCGACACGATGACTACACTACCTACCCGAGACGGGGGATGTGGAGCGCTCTACGACATC 10431
QY      379  tggttcagtcgccgtcacgaattccgcgcgaacagcgtacagcgcgcgcgcgcgcgcgcgtgatac 438
Db 10432  TGCCTGGATTTCACACACCACAGACGACCGGGGTCAACACAGCA-----GGAGATCATGTATC 10485
QY      439  tggctgaactgaaacgggcgcggtgagtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 498
Db 10486  TGGTTCAACCAACCAAGGCGTCCATTCAGCCGGTCCGCTCCCGGTGGGCAACCAACCAATC 10545
QY      499  gccgcggccacacttggaagctctgatactgcgcgaacttggaactggaattacaatcgcctaccgg 558
Db 10546  GAGGGCAAGAACTTGTTGGTGTGGGTGAGACCAACCGCATGACACGCGATGAGCTAT 10605
QY      559  cgcgaagacgcccaacacgctcggttgagcgaactgcgaacccgaaagcccttcatcgaaagcgcg 618
Db 10606  GTGCGGACCGAGACCATCGAGGTCTGAGACTTCGACGCTGATGAGTTGTGTCACCAACACC 10665
QY      619  gtcgcgcgcgcgtacataccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 678
Db 10666  GCCACCATGAGACCGCATACCGACTCTGTGTACTCTACAGAGAGATCCGGGGCGCGCTTTGGAG 10725
QY      679  ctctctggagagcggggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 728
Db 10726  CCCTGGAGCGAGGTGTGGGTGTGGGGGTGATGCTCTCGCGCAAGT 10775

Search completed: August 19, 2002, 12:45:06
Job time: 10450 sec

```

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 13:53:47 ; Search time 544 Seconds  
(without alignments)  
2310.260 Million cell updates/sec

Title: US-10-003-759-3\_COPY\_52\_783

Perfect score: 1 tgcgactggtcttcccca.....atttccgaacggtgcag 732

Sequence: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Scoring table: 1736436 segs, 858457221 residues

Searched: Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq.032802:\*

1: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.8	18.0	1116	20	AAV72359
2	131.8	18.0	1116	21	AAZ57029
3	131.6	18.0	1715	21	AAZ57031
4	57.6	7.9	1983	22	AAA54513
5	53.6	7.3	77536	21	AAA14651
6	49.8	6.8	756	17	AAT16766
7	49.4	6.7	1404	21	AAA38763
8	49	6.7	756	14	AAO43032
9	48.8	6.7	985	19	AAV64548

10	48.8	6.7	985	19	AAV44439	Mycobacterium tube
11	48.8	6.7	985	20	AAZ19349	M. tuberculosis an
12	48.8	6.7	985	20	AAZ19137	M. tuberculosis re
13	48.4	6.6	12152	22	AAZ08699	M. carbonacea DNA
14	48	6.6	4403765	22	AA199683	Mycobacterium tube
15	47.8	6.5	24379	18	AAT93095	Streptomyces freno
16	47.8	6.5	24379	18	AAV25925	Streptomyces roseo
17	47.6	6.5	30001	19	AAT61016	Total DNA sequence
18	47.6	6.5	30001	20	AAZ05110	S. aureofaciens DN
19	47.2	6.4	1908	16	AAO93364	S. lividans protea
20	47.2	6.4	1908	20	AAV84065	Tripeptidyl aminop
21	47.2	6.4	1908	21	AAZ61403	cDNA sequence enco
22	46.8	6.4	15872	21	AAZ61283	S. venezuelae vep
23	46.8	6.4	20394	22	AAZ24892	Pimaricin biosynth
24	46	6.3	44377	18	AAT78508	Platanolide synth
25	46	6.3	44377	18	AAT80414	Myceliophthora the
26	45.6	6.2	744	18	AAT74073	Mycobacterium tube
27	45.2	6.2	1291	20	AAZ87940	M. tuberculosis im
28	45.2	6.2	2852	19	AAV64558	Mycobacterium tube
29	45.2	6.2	2852	19	AAV44449	M. tuberculosis an
30	45.2	6.2	2852	20	AAZ19359	M. tuberculosis re
31	45.2	6.2	2852	20	AAZ19147	Nucleotide sequenc
32	45.2	6.2	77536	21	AAA14651	Non-reducing sacch
33	44.8	6.1	2268	21	AAA10501	Rice 1-deoxy-D-xy
34	44.8	6.1	2618	21	AAA38760	Non-reducing sacch
35	44.8	6.1	3252	21	AAA10505	Mouse Sox1 cDNA.
36	44.6	6.1	2376	20	AAZ16153	Polyanion brachys
37	44.6	6.1	2712	19	AAT94214	L05390 cDNA clone.
38	44.6	6.1	5224	22	AAZ90079	S. venezuelae macr
39	44.6	6.1	11220	21	AAZ87298	Sequence encoding
40	44.6	6.1	12588	15	AAZ063293	S. venezuelae pik
41	44.6	6.1	36778	21	AAZ87318	Nucleotide sequenc
42	44.6	6.1	37948	21	AAZ87285	Recombinant cosmid
43	44.6	6.1	38506	21	AAZ75633	H. virescens inorg
44	44.6	6.1	38506	21	AAZ56001	
45	44.4	6.1	867	22	AAH48828	

#### ALIGNMENTS

RESULT 1	AAV72359	standard; DNA; 1116 BP.
ID	AAV72359	
XX	AAV72359;	
AC		
XX		
DT	28-JUL-1999	(first entry)
XX		
DE	Actinomyces sp. 36kd cellulase DNA.	
XX		
XX	Cellulase; detergent; animal feed; nutritional value; textile;	
KW	stone washing; texture modification; appearance; cellulosic fabric;	
KW	pulp; draining; paper; baking additive; starch treatment; grain;	
KW	high-fructose corn syrup production; ethanol production; fibre reduction;	
KW	milling; ss.	
XX		
OS	Actinomyces sp.	
XX		
PN	WO9925846-A2.	
PD		
XX	27-MAY-1999.	
PF		
XX	18-NOV-1998;	98WO-US24649.
PR	24-JUN-1998;	98US-0104308.
PR	19-NOV-1997;	97US-0974041.
XX	19-NOV-1997;	97US-0974042.
XX		
PA	(GENV ) GENENCOR INT INC.	
PI		
XX	Jones BF, Van Der Kleij WAH, Van Solingen P, Weyler W;	

DR WPI: 1999-347481/29.  
DR P-PSDB: AA084473.  
XX  
XX  
PT New Actinomycete cellulase useful in detergent compositions, in  
PT animal feeds and in treatment of textiles  
XX  
XX  
PS Claim 4: Fig 2; 36pp; English.  
XX  
XX This invention describes a novel cellulase isolated from an Actinomycete  
CC sp. which can be used in detergent compositions, as animal feeds (to  
CC increase nutritional value) and in treatment of textiles (e.g. stone  
CC washing or modifying texture, feel and/or appearance of cellulosic  
CC fabrics, including removal of 'immature' or 'dead' cotton) pulp (to  
CC improve draining) and paper. They may also be used as baking additives,  
CC for treating starch (in production of high-fructose corn syrup or  
CC ethanol) and for treating grain (to reduce fibre during milling).  
XX  
XX Sequence 1116 BP; 191 A; 432 C; 340 G; 153 T; 0 other;

Query Match	18.0%;	Score 131.8;	DB 20;	Length 1116;
Best Local Similarity	50.7%;	Pred. No. 4e-18;		
Matches 347;	Conservative	0;	Mismatches 332;	Indels 6; Gaps 1

[illegible]

```

XX RESULT 2
XX ID AA257029 standard; DNA; 1116 BP.
XX AC AA257029;
XX DT 19-MAY-2000 (first entry)
XX DE Actinomycetes cellulase protein encoding DNA.
XX KW Cellulase; Actinomycetes; detergent; feed additive; textile treatment;
XX KW pulp; paper; ds.
XX OS Streptomyces sp.
XX Key Location/Qualifiers
XX CDS 1..1116
XX FT /*tag= a
XX FT /product= "cellulase"
XX FT sig_peptide 1..81
XX FT /*tag= b
XX FT mat_peptide 82..1113
XX FT /*tag= c
XX WO200009707-A1.
XX PD 24-FEB-2000.
XX PF 28-MAY-1999; 99WO-US11971.
XX PR 24-JUN-1998; 98US-0104308.
XX PR 18-NOV-1998; 98WO-US24649.
XX PR 28-MAY-1999; 99US-0321981.
XX PA (GEMV ) GENENCOR INT INC.
XX PI Jones BE, Van Der Kleij WAH, Van Solingen P, Weyler W;
XX DR WPI; 2000-224344/19.
XX DR P-PSDB; MAY67496.
XX PT A novel Actinomycetes cellulase and related DNA, useful for detergent
XX FT compositions, treating textiles and paper or pulp
XX PS Claim 5; Fig 2; 72pp; English.
XX The invention provides a cellulase from Actinomycetes. The cellulase can
XX be used in a detergent composition, as an additive for animal feed and
XX for the treatment of textiles or pulp and paper. The DNA encoding the
XX cellulase can be used to identify homologous cellulases and for
XX recombinant production of cellulases. The present sequence represents a
XX DNA encoding a cellulase from Actinomycetes.
XX Sequence 1116 BP; 191 A; 432 C; 340 G; 153 T; 0 other;

```

Query Match	18.0%;	Score 131.8;	DB 21;	Length 1116;
Best Local Similarity	50.7%;	Pred. No. 4e-18;		
Matches 347;	Conservative 0;	Mismatches 332;	Indels 6;	Gaps 1;

[illegible]







```

FT CDS complement (12113..13150)
FT /tag= l
FT /note= "fkbl gene"
FT CDS complement (13212..23988)
FT /tag= m
FT /note= "fkbc gene"
FT misc-feature complement (13452..13662)
FT /tag= n
FT /note= "ACP6"
FT misc-feature complement (13761..14394)
FT /tag= o
FT /note= "KR6"
FT misc-feature complement (14517..15294)
FT /tag= p
FT /note= "ER6"
FT misc-feature complement (15438..16587)
FT /tag= q
FT /note= "dehydratase domain (DH) 6"
FT misc-feature complement (16587..17820)
FT /tag= r
FT /note= "acyltransferase domain (AT) 6"
FT misc-feature complement (17820..19053)
FT /tag= s
FT /note= "KS6"
FT misc-feature complement (19116..19326)
FT /tag= t
FT /note= "ACPS"
FT misc-feature complement (19464..20097)
FT /tag= u
FT /note= "KR5"
FT misc-feature complement (20241..21420)
FT /tag= v
FT /note= "DH5"
FT misc-feature complement (21420..22653)
FT /tag= w
FT /note= "AT5"
FT misc-feature complement (22653..23892)
FT /tag= x
FT /note= "KS5"
FT CDS complement (23992..46573)
FT /tag= y
FT /note= "fkpb gene"
FT misc-feature complement (24163..24373)
FT /tag= z
FT /note= "ACP4"
FT misc-feature complement (24997..26146)
FT /tag= aa
FT /note= "DH4 (inactive)"
FT misc-feature complement (26146..27430)
FT /tag= ab
FT /note= "AT4"
FT misc-feature complement (27430..28684)
FT /tag= ac
FT /note= "KS4"
FT misc-feature complement (28750..28960)
FT /tag= ad
FT /note= "ACP3"
FT misc-feature complement (29092..29740)
FT /tag= ae
FT /note= "KR3"
FT misc-feature complement (29869..31018)
FT /tag= af
FT /note= "DH3 (inactive)"
FT misc-feature complement (31018..32185)
FT /tag= ag
FT /note= "AT3"
FT misc-feature complement (32185..33439)
FT /tag= ah
FT /note= "KS3"
FT misc-feature complement (33505..33715)
FT /tag= ai
FT /note= "ACP2"
FT misc-feature complement (33823..34480)

```

---

```

FT /tag= aj
FT /note= "KR2"
FT misc-feature complement (34606..35749)
FT /tag= ak
FT /note= "DH2 (inactive)"
FT misc-feature complement (35749..37144)
FT /tag= am
FT /note= "AT2"
FT misc-feature complement (37145..38296)
FT /tag= an
FT /note= "KS2"
FT misc-feature complement (38371..38581)
FT /tag= ao
FT /note= "ACP1"
FT misc-feature complement (38677..39307)
FT /tag= ap
FT /note= "KR1"
FT misc-feature complement (39442..40609)
FT /tag= aq
FT /note= "DH1"
FT misc-feature complement (40609..41842)
FT /tag= ar
FT /note= "AT1"
FT misc-feature complement (41842..43093)
FT /tag= as
FT /note= "KS of extender module 1 (KS1)"
FT misc-feature complement (43144..43660)
FT /tag= at
FT /note= "ACP of loading domain"
FT misc-feature complement (43777..44629)
FT /tag= au
FT /note= "ER of loading domain"
FT misc-feature complement (44974..46573)
FT /tag= av
FT /note= "CoA ligase of loading domain"
FT CDS 46754..47788
FT /tag= aw
FT /note= "fkpo gene"
FT CDS 47785..52272
FT /tag= ax
FT /note= "fkdp gene"
FT CDS 52275..71465
FT /tag= ay
FT /note= "fkpa gene"
FT misc-feature 52362..53576
FT /tag= az
FT /note= "KS7"
FT misc-feature 53577..54716
FT /tag= ba
FT /note= "DH7"
FT misc-feature 54717..55871
FT /tag= bb
FT /note= "ER7"
FT misc-feature 56943..57575
FT /tag= bd
FT /note= "KR7"
FT misc-feature 57710..57920
FT /tag= be
FT /note= "ACP7"
FT misc-feature 57990..59243
FT /tag= bf
FT /note= "KS8"
FT misc-feature 59244..60398
FT /tag= bg
FT /note= "AT8"
FT misc-feature 60399..61412
FT /tag= bh
FT /note= "DH8 (inactive)"

```



```

FT      /tag- a
FT      /product- "1-deoxy-D-xylose 5-phosphate synthase"
FT      /partial
PN      WO200032792-A2.
XX      08-JUN-2000.
XX      02-DEC-1999; 99WO-US28587.
XX      03-DEC-1998; 98US-0110779.
XX      (DUPO ) DU PONT DE NEMOURS & CO E. I.
PI      Cahoon RE, Tao Y, Williams ME, Coughlan SJ, Weng Z;
DR      MPI; 2000-412338/35.
DR      P-PSDB; AAY97425.
XX      Polynucleotide encoding 1-deoxy-D-xylose 5-phosphate synthase enzyme
PT      useful for producing transgenic plants and for producing antibodies
PT      specific to which is useful for screening cDNA expression libraries -
XX      Claim 2; Page 67; 73pp; English.
XX      The present sequence is a putative coding sequence for the wheat
CC      1-deoxy-D-xylose 5-phosphate synthase enzyme (DXPS). Its protein is
CC      involved in the isoprenoid biosynthesis pathway. The cDNA was identified
CC      by sequencing a number of clones and then comparing their protein
CC      sequences to known proteins: this showed the sequence's similarity to the
CC      Capsicum annuum DXPS sequence. The DXPS gene and protein can be used to
CC      create transgenic plants which express the gene at either different
CC      levels or at different stages of development compared to normal, and to
CC      identify herbicides.
XX      Sequence 1404 BP; 277 A; 482 C; 439 G; 206 T; 0 other;
SQ
Query Match      6.7%; Score 49.4; DB 21; Length 1404;
Best Local Similarity 46.9%; Pred. No. 0.23;
Matches 220; Conservative 0; Mismatches 246; Indels 3; Gaps 2;
QY      178 ggcgaattccgatacagcgccgatacagacaagcaacgctggccgtactcgcg 237
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      355 ggcgtcgtcagatcagctcagcgcgtcgtcgtcgtcgtcgtcgtcgtcgtc 414
QY      238 ggcattactcgtggctggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 297
DB      | || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      415 cttctggagctggccaccgctcgtacccgacaagaattcgaaggcgcgcgataag 474
QY      298 gtgcagagagctgtcgcagctgtcgcagagctgtcgcagctcgcagatcagacgagcgc 357
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      475 atgcgcagcgtg--cgcgcagaccacgcgtgtcgcgttcgtcaagcgctcgcag- agc 531
QY      358 tggaaatgcgcgtcagcaatctgttcaatccgtcagcaatccgcgcgcgtacagc 417
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      532 gaattacacagcttcgtcgcagcagcagcagcagcagcagcagcagcagcagcagc 591
QY      418 gggcgcccgagctgtatgactgtgaactggaagcgcgcgctgtatggcgcgcgagc 477
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      592 ggcgttcggggggaacctcaaggcgcggaagaacagctgtgtgggtgatttgggagcg 651
QY      478 cgcgttgcacacgtgtgaactgtgcggggcgccactgtggaagtctgtatgtcgcagctggagc 537
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      652 gcatgacgagcgcgcgcggtacgagcgatgaacaacgcgctacccctcgcagctggac 711
QY      538 tggaaatgcgtcgtacacgcgcgagcagcagcagcagcagcagcagcagcagcagcagc 597
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      712 atgactgtatctcctcaacgaacaaagcagctgtcgtcgcagcagcgagcagcagcagcagc 771
QY      598 aagcgcttcatcagcagcagcgttcggcgcgctacacccgcgcgcgcgcgcgcgcgcgcgcgc 646
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      772 ccggcgccgcgcgcgtgtggcgcgctcagcgcgccctcagcaagcagcagcagcagcagcagc 820

```

```

RESULT      8
AAQ43032
ID      AAQ43032 standard; DNA; 756 BP.
XX
AC      AAQ43032;
XX
DT      07-SEP-1993 (first entry)
XX
DE      Collagen-like polymer DCP3-C2(AB12)C2 coding sequence.
XX
KW      Recombinant; collagen-like polymer; CLP; tripeptide; helix; membrane;
KW      fibre; film; coating; triad sequence; collagen; mammalian; moulding;
KW      hydrogel; interchain linkage; colloid suspension; DCP; antibody; ds.
XX
OS      Synthetic.
XX
PN      WO9310154-A.
XX
PD      27-MAY-1993.
XX
PF      04-NOV-1992; 92WO-US09485.
XX
PR      12-NOV-1991; 91US-0791960.
XX
PA      (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
XX
PI      Cappello J, Ferrari FA;
XX
DR      MPI; 1993-182496/22.
DR      P-PSDB; AAR37738.
XX
PT      High mol. wt. collagen-like protein polymers - capable of being
PT      produced in unicellular microorganisms
XX
PS      Disclosure; Page 39; 82pp; English.
XX
CC      The sequences given in AAQ43030-34 encode examples of recombinantly
CC      produced DCP collagen-like polymers (CLPs) which consist of repeated
CC      tripeptide sequences selected from a wide range of GXY sequences.
CC      where X and Y can be any amino acid. These sequences can be cloned
CC      into plasmids and used to transform E. coli to produce the DCP
CC      proteins. DCP peptides comprise repeated units of: A = GAPGAGP,
CC      B = GSRGDPGP and/or C = GAHGAPGPK. These polymers may be used to
CC      raise anti-DCP antibodies in rabbits. These polymers have molecular
CC      weights of >30 kD and are able to form helices due to interchain
CC      linkages. These polymers pref. contain a proportion of tripeptide
CC      triad sequences found in natural collagens, pref. mammalian collagens.
CC      The CLPs impart unique characteristics to materials such as fibres,
CC      membranes, films, coatings, hydrogels, colloid suspensions and moulded
CC      articles.
XX
SQ      Sequence 756 BP; 84 A; 288 C; 291 G; 93 T; 0 other;
Query Match      6.7%; Score 49; DB 14; Length 756;
Best Local Similarity 43.5%; Pred. No. 0.27;
Matches 223; Conservative 0; Mismatches 290; Indels 0; Gaps 0;
QY      72 gtgcgagcgtgtgagcgcgcgagctgtgcccggggcgctaccgggtgatacaaacgt 131
DB      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      218 gtgcgctgtgagccggtgtgtccacacggtgtctcgggacctgagcccgccaggtgcgc 277
QY      132 atggggcgcgagagaccgcccagtgcatgtgaggtgcgcgaacttgaagcggaactaacgat 191
DB      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      278 ctggaccggtctgttccacgcgggtgtcctcgggaaccttgaggcccgccaggtgtgcgcggagc 337
QY      192 caacggcgcatcagcagcagcagcaacacgtgtgcgcgcgtactacggtctactactcgg 251
DB      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      338 cgcgtgtgtccacacgggtgtgtcgcgggaaccttcagcccgccaggtgtgcgcgtgcgcgcgc 397
QY      252 gtgcactgtggcgccctgtcagcagcaattcggattgtccggcgcgctgtcaggaagcgtgtc 311

```



PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and diagnosis of tuberculosis

PS Claim 4; Page 187-188; 250pp; English.

XX This is the 3' region of DNA coding for an antigenic portion of  
CC Mycobacterium tuberculosis antigen XP25; 5' DNA is provided in  
CC AA44438.. XP25 DNA was isolated from a M. tuberculosis strain Erdman  
CC genomic DNA expression library using sera from patients having  
CC extrapulmonary tuberculosis. It bears no similarity to known  
CC sequences. The invention relates to methods and compositions for  
CC diagnosing tuberculosis. It provides polypeptides (see  
CC AA44438.. XP25) comprising an antigenic portion of a M.  
CC Mycobacterium tuberculosis antigen, as well as DNA sequences encoding such  
CC polypeptides, recombinant expression vectors and transformed or  
CC transfected host cells. Also claimed are methods and diagnostic  
CC kits for detecting M. tuberculosis infection in a patient using  
CC these polypeptides, antibodies or oligonucleotide probes and  
CC primers.

SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 6.7%; Score 48.8; DB 19; Length 985;  
Best Local Similarity 44.9%; Pred. No. 0.3;  
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

OY 13 ttcccgatgagcgaacgaagaaagacgagctgagcccgagccgagctgagctg 72  
DB 175 ttcccgatgagcgaacgaagaaagacgagctgagcccgagccgagctgagctg 234  
OY 73 tgcgagcgtgagcgcgcgagctgagccgagccgagctgagccgagctgagccg 132  
DB 235 accggtacgagcgcgcgagctgagccgagccgagctgagccgagctgagccg 294  
OY 133 tgggagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 192  
DB 295 gccgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 354  
OY 193 acaagcgcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 252  
DB 355 ggccttgaagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 414  
OY 253 tgcgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 308  
DB 415 atcaagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 474  
OY 309 gtcgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 368  
DB 475 gccgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 533  
OY 369 ctgaagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 428  
DB 534 cggagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 593  
OY 429 gctgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 488  
DB 594 gggagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 653  
OY 489 cgtgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 548  
DB 654 cggagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 713  
OY 549 cgcctacgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 602  
DB 714 cacttgatcgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 767

RESULT 11  
AA219349  
ID AA219349 standard; DNA: 985 BP.

XX AA219349;  
AC  
XX  
XX 05-NOV-1999 (first entry)  
DE  
XX  
XX M. tuberculosis antigen 3' XP25 DNA sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
XX immunotherapy; diagnosis; immunisation; vaccine; infection;  
XX immune response; skin test; ss.

OS Mycobacterium tuberculosis.

XX WO942076-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03268.

XX 05-MAY-1998; 98US-0072967.

XX 18-FEB-1998; 98US-0025197.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skelky YAW, Twardzik DR, Vedvick TS;  
XX WPI; 1999-527409/44.

XX

PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
PT skin tests and protective or therapeutic vaccines or compositions

PS Claim 4; Page 171-172; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to  
CC AA219225 are used in the exemplification of the present invention.

SQ Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

Query Match 6.7%; Score 48.8; DB 20; Length 985;  
Best Local Similarity 44.9%; Pred. No. 0.3;  
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

OY 13 ttcccgatgagcgaacgaagaaagacgagctgagcccgagccgagctgagctg 72  
DB 175 ttcccgatgagcgaacgaagaaagacgagctgagcccgagccgagctgagctg 234  
OY 73 tgcgagcgtgagcgcgcgagctgagccgagccgagctgagccgagctgagccg 132  
DB 295 gccgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 354  
OY 133 tgggagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 192  
DB 295 gccgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 252  
OY 193 acaagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 252  
DB 355 ggccttgaagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 414  
OY 253 tgcgagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 308  
DB 415 atcaagcgcgcgagcgcgcgagctgagccgagccgagctgagccgagctgagcc 474

QY 309 gtccgacgtgctgacgagctggaagctcaacgctacgacgagccgctggaatgcgc 368  
DB 475 gccgcgaagctcgcgacaacggcggtgcggctgacggtggtgc-cgltggcgcgcc 533  
QY 369 ctacgaatctgtctcaagctccgaattccggacaagcgctacacggcgccgca 428  
DB 534 cggcaacggcgacaaacggcggtcgcctgacacgacccaaggccggcgacggcgccgc 593  
QY 429 gctgatatctgtgctgaactggaacggcggtgatatgctggcgcgacccgctggccac 488  
DB 594 gggcaatgvcggaacggcggtgcgcggctgctgvcggtgcggcgacacaaatttcaa 653  
QY 489 cggggaactggtggcgccacccctgggaagtctgtatgctggacttgggaattcat 548  
DB 654 cggcgccgaaggtgtgtgctggcgccaaaggcgccaaaggcggtcttggcggaagcac 713  
QY 549 cgactacggcgacgacgacccacacgtctggtgagcggaactggaactggaagc 602  
DB 714 cactgatatcgctagcgcgaccccggaagaagccgattccaaacaggcgacgattgcc 767

## RESULT 12

AA219137  
ID AA219137 standard; DNA; 985 BP.

XX AA219137;

XX 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen DNA encoding 3' XP25.

KM Antigen; diagnosis; detection; infection; antibody; immunisation;  
KW vaccine; immunity; ss.

XX Mycobacterium tuberculosis.

XX MO942118-A2.

PN 26-AUG-1999.

PF 17-FEB-1999; 99MO-US03265.

XX 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes WD, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;  
XX WPI; 1999-527416/44.

PT New polypeptide comprising antigenic portions of M. tuberculosis

XX Claim 4; Page 217; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding  
CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
CC polypeptides are useful for detecting M. tuberculosis infection in a  
CC biological sample by detecting antibodies which bind with the  
CC polypeptides, and are useful as vaccines for immunizing against  
CC M. tuberculosis infection. The new detection methods are needed as  
CC current vaccination strategies do not provide 100% immunity.

XX Sequence 985 BP; 138 A; 330 C; 423 G; 94 T; 0 other;

XX Query Match 6 7%; Score 48.8; DB 20; Length 985;

XX Best Local Similarity 44.9%; Pred.No. 0.3; Mismatches 322; Indels 5; Gaps 2;

QY 13 ttcccgatcgcgacacggaagacggacggtcgacccgacgacgctcgagctg 72  
|| ||| ||||| | ||| | | | | | | | | | | | | | | | | | | |

DB 175 ttcccggtgtgctgcggcgacgaagtgtgcggtgcccggcggtggcggtggcg 234  
QY 73 tgcggaacgtctggagacgcgcgcgaatgtgtgcccggggggtgtaaccaaacgta 132  
DB 235 accggtatccggcgacacccggcggtgtgtgtggtggccacccgtatgtgacgtccgg 294  
QY 133 tgggtgcggaagacgcgccgaatgtcatgtgaagcttgaaacgggaacttcacgact 192  
DB 295 gccggcggtgcgcgcgtgtgacggtgcggtgtggtggccagcggtctcggtggcccttc 354  
QY 193 acacgggctgatacgaacaaacggtgtgcccgtatccgacatccgacatcttcgcgg 252  
DB 355 gacttgaacggcgacgaagcggtgacgaagggcggtgcgacggtgcgcggcgccg 414  
QY 253 tgcgaactgggtgcgctgtcgaagc---aatcgggatttgcggtgcggtgcaagagct 308  
DB 415 atcaacggggcggtgcggtgcgcgcgcgacacggcggtgcggtggaacggcgaacggt 474  
QY 309 gtccgacgtgtgcgacgagcttgacgctcaacgctacgaacggcggtctggaatgcgc 368  
DB 475 gccgcaggtctcgcgacacacggcggtgcggtgcgtgacggtgtgtggtgc-cgtgtg 533  
QY 369 ctacgaatctgtgctgaatccgctcaacgaattccggaacgggtgacggtgcgcgca 428  
DB 534 cggcaacggcggaacacggcggtgcgcgtgacgaacgaagccggcgacggtgcgcgcgc 593  
QY 429 gctgatatctgtgctgaactggaacggcggtgatatgctggacttgggaattcat 488  
DB 594 gggcaatgvcggaacggcggtgcgcggctgctgvcggtgcggcgacacaaatttcaa 653  
QY 489 cgttgaactggcggtggccaacttgggaagtctgtatgctggacttgggaattcat 548  
DB 654 cggcgccgaaggtgtgtgctggcgccaaaggcgccaaaggcggtcttggcggaagcac 713  
QY 549 cgactacggcgacgacgacccacacgtctggtgagcggaactggaactggaagc 602  
DB 714 cactgatatcgctagcgcgaccccggaagaagccgattccaaacaggcgacgattgcc 767

## RESULT 13

AA508699/c  
ID AA508699 standard; DNA; 12152 BP.

XX AA508699;

XX 26-SEP-2001 (first entry)

DE M. carbonacea DNA encoding Everninomycin biosynthetic enzymes.

XX Everninomycin; antibiotic; bottle-neck gene; orthomycin;  
KW fermentation; ds.

XX Micromonospora carbonacea var. africana.

XX Key location/Qualifiers

FT CDS complement (189..1064)

FT /\*tag- a

FT /product- "ORF1"

FT complement (1069..1073)

FT /\*tag- b

FT 1184..2767

FT /product- "ORF2"

FT complement (2683..3753)

FT /\*tag- d

FT complement (3766..4276)

FT /\*tag- e

FT /product- "ORF4"

FT complement (4280..4284)

FT /\*tag- f

FT complement (4526..5368)

FT /\*tag- g



FT FT /product= "ORF5"  
FT CDS complement (5392..6147)  
FT /tag= h  
FT /product= "ORF6"  
FT RBS complement (6152..6156)  
FT /tag= i  
FT CDS complement (6194..7282)  
FT /tag= j  
FT /product= "ORF7"  
FT CDS complement (7280..8133)  
FT /tag= k  
FT /product= "ORF8"  
FT RBS complement (8141..8145)  
FT /tag= l  
FT CDS complement (8254..9318)  
FT /tag= m  
FT /product= "ORF9"  
FT RBS complement (9324..9328)  
FT /tag= n  
FT CDS 9575..10504  
FT /tag= o  
FT /product= "ORF10"  
FT RBS 9568..9571  
FT /tag= p  
FT CDS 10584..11585  
FT /tag= q  
FT /product= "ORF11"

PN WO200151639-A2.  
XX  
XX 19-JUL-2001.  
PD  
PF 12-JAN-2001; 2001WO-US01187.  
XX  
XX 12-JAN-2000; 2000US-O175751.  
PR  
XX  
PA (SCHE ) SCHERING CORP.  
PI Hosted TJ, Horan AC, Wang TX:  
PI  
XX  
DR MPI: 2001-442147/47.  
DR P-FSDB; AAU04901, AAU04902, AAU04903, AAU04904, AAU04905, AAU04906,  
DR AAU04907, AAU04908, AAU04909, AAU04910, AAU04911.  
XX  
XX New nucleic acid molecules encoding evernimicin pathway gene  
PT products, useful for improving yields of evernimicin, to produce new  
PT evernimicin and as probes to identify homologous sequences -  
PS  
XX

Example 1; Fig 12; 109pp: English.

The sequence encodes 11 proteins comprising enzymes of the  
CC evernimicin antibiotic biosynthetic pathway. A vector comprising a  
CC M. carbonacea evernimicin biosynthetic pathway resistance gene product  
CC is useful for selecting for a transfected or transformed host cell. An  
CC integrative version of the vector is useful for introducing a  
CC evernimicin pathway gene (a bottle-neck gene) into an actinomycete of  
CC the genus Micromonospora. The DNA encoding the biosynthetic proteins is  
CC useful for synthesising novel evernimicin-related compounds, arising  
CC from modifications of the DNA sequence designed to change glycosyl and  
CC modified orsellinic acid groups contained in evernimicin, for  
CC expressing functional or mutant evernimicin biosynthetic enzyme for  
CC evaluation, diagnosis and preferably biosynthesis of evernimicin or  
CC other secondary metabolic products, improving the yield of evernimicins  
CC to produce novel evernimicins and also as a hybridisation probe to  
CC identify homologous sequences. The encoded polypeptides are useful for  
CC combinatorial biosynthesis to generate libraries of orthomycins, e.g.,  
CC evernimicin analogues/homologues and drug discovery. The  
CC DNA encoding the integrase allows for increasing a given gene dosage. The  
CC integrative vector can be used to permanently integrate copies of a  
CC heterologous gene of choice into chromosomes of different hosts and to  
CC integrate genes which increase the yield of known products or to generate  
CC novel products such as hybrid antibiotics or other novel secondary  
CC metabolites. The vector can also be used to integrate antibiotic

CC	resistance genes in order to carry out bioconversions with compounds to
CC	which the strain is normally sensitive and is thus useful in fermentation
CC	processes involving e.g. Streptomyces antibioticus.
xx	
SQ	Sequence 12152 BP; 1675 A; 4502 C; 4475 G; 1500 T; 0 other;
	Query Match            6.6%; Score 48.4; DB 22; Length 12152;
	Best Local Similarity 45.9%; Pred. No. 0.4;
	Matches 166; Conservative 0; Mismatches 196; Indels 0; Gaps 0;
OY	287 tgcgcgagcgcgctgcaaggactgtccgaacgttgcacagcagtggacgctcacgccgatca 346 
Db	8021 TGTCGGCGCAACTCGCATGTGACACCCGAGAGACCGGAGACAACCTGTCCGCCGAGTGTCTC 7962
OY	347 cgacggcgccgcctggaaatgcgcgcctcagacatcttgattcaagtcctcagaattccgaca 406 
Db	7961 GGCTGCGCTCCAGAGACCCTCGAGCCCTTCGTGTGCGAGCGCACCCAGCCGCCGCGCTGCACG 7902
OY	407 acgggtacacgc 466 
Db	7901 CGCGCTTTGCAGCAGCTGCTCGGGGGTGGGGCGCTGGAAAGCCGTTGGAACCGGATCGGCACCT 7842
OY	467 cgggc 526 
Db	7841 TCCCGGTCGCGGCTTCCCGGGGCCAAGCCGCCGAGACACGAGAGATACGAGCTGGCACCA 7782
OY	527 ccgaactggagactggagattacatcgctcctaaccgcgcacgacgcacacacacgttcgtgagcg 586 
Db	7781 TCGAGCGCCAGTTTCTGTGGCCGACGACGCGCACGGGCTGTGGCCCCAGAACTGGGAGGGCG 7722
OY	587 agactgacctgaaggcctcatcgacgc 646 
Db	7721 AGCTGGACCTGCTGCCGCCCACTACGCGGAAGTCTTCCGCTGCAACCTCCGCTGGCGGG 7662
OY	647 gg 648 
Db	7661 GG 7660
	RESULT 14
ID	AA199683/C
ID	AA199683 standard; DNA; 4403765 BP.
XX	
AC	AA199683;
XX	
DT	15-JAN-2002 (first entry)
XX	
DE	Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX	
KM	Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome; variation; epidemiology; patient treatment; epidemic monitoring; ds.
OS	Mycobacterium tuberculosis.
PN	US6294328-B1.
XX	
PD	25-SEP-2001.
XX	
PF	24-JUN-1998; 98US-0103840.
XX	
PR	24-JUN-1998; 98US-0103840.
XX	
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Fleischmann RD, White OR, Fraser CM, Venter JC;
XX	
DR	WPI, 2001-647261/74.
XX	
PT	Evaluating strain variation of Mycobacterium tuberculosis, comprises determining the nucleotide sequence of the strain at positions in the genome corresponding to positions found in M. tuberculosis strains CDC 1551 and H37Rv differ -

XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
 PS  
 CC The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen.  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (A199683) and  
 CC H37Rv (A199682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?docID=629432881.  
 CC  
 XX  
 SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 6.6%; Score 48; DB 22; Length 4403765;  
 Best Local Similarity 45.6%; Pred. No. 0.62;  
 Matches 208; Conservative 0; Mismatches 245; Indels 3; Gaps 1;

QY 100 gccgggggagcgctaccgggtgatcaacaacgtatggggcgagagccgcagtcgcaat 159  
 DB 3963782 GTCCGCAACGGGACCTCGGCAACCACTGGGAGCGGCAATACCGCAACTCCAAC 3963723  
 QY 160 gaagtcgagctggaagcagcgcaactcagatcacagcgccgcatcacagcgcaac 219  
 DB 3963722 GTCCGGGTTCCGGAAATACCGGCANACAGTCCGGCACCGGAAACGCCGCGCAAC 3963663  
 QY 220 aacgtgacgcctatccggccatctacttcgggtgccaatggggcgccgtgacagcaat 279  
 DB 3963662 ATCCGCGCGCGGAGACCGCGAGCAACCTGGGGCTTCGGCAACACGCAATGGCAAC 3963603  
 QY 280 tcgggattccgcgcgcgtcgcag---agctgtccgaagtgcgcagagctgacgcctc 336  
 DB 3963602 ATCCGTTTCGGGACACCGCGCAACCACTCGGCTTCGGGCTCACCGCAACCAACCG 3963543  
 QY 337 acgcgcatacagcagcgccgctgaaatgcgcctacagacatctggttcagttccgcgtcaag 396  
 DB 3963542 GTGGGCAATCGGTGCTGTAACCTCGGCGACGGAACATCGCTTGTTCACCTCGGGCAC 3963483  
 QY 397 aattccggaacgcctacagcgcgccgcgcagctgatatgctgctgacatggaacgyc 456  
 DB 3963482 AACCAACGTGCTTCTTCAACTCCGCAACCTGGGTATCGGCAACTCCAGCGAC 3963423  
 QY 457 ggcgtatgcgcggcgccgcgcgcgtgcgcacgctggaactgycgcgggaccacctgggaa 516  
 DB 3963422 GCGAAATGGGCACTCGGAACCTCGGGGCCGCTGCTCGCGGGCCATTAAC 3963363  
 QY 517 gtctgtatgcgcagctggaacttaacatgcgc 552  
 DB 3963362 ACCGGTTTCGGGAACCTCGGCTTCGCTCAACACGGGC 3963327

RESULT 15  
 AAT93095  
 ID AAT93095 standard; cDNA; 24379 BP.  
 XX  
 AC AAT93095;

XX 11-MAY-1998 (first entry)  
 XX  
 DE Streptomyces frenolicin gene cluster.  
 XX  
 KW Frenolicin; antibiotic; feed additive; anticoccidial;  
 KW coccidiostatic; efflux pump; butyrate starter synthase;  
 KW polyketide synthase; PKS; hemiketalase; ketoreductase; cyclase;  
 KW dehydrase; ketoreductase; hydroxylase; Streptomyces roseofulvus;  
 KW ds.  
 XX

OS	Streptomyces sp.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	636..2948
FT		/*tag= a
FT		/product= 80 kDa non-membrane protein
FT		/note= "gene A (specifically claimed)"
FT	CDS	2945..3916
FT		/*tag= b
FT		/product= membrane protein
FT		/note= "gene B (specifically claimed)"
FT	CDS	4020..4844
FT		/*tag= c
FT		/product= protein with 6 membrane-spanning domains
FT		/note= "gene C (specifically claimed)"
FT	CDS	4841..6415
FT		/*tag= d
FT		/product= ATP-binding component of ABC transporter
FT		/note= "gene D (specifically claimed)"
FT	CDS	6533..7183
FT		/*tag= e
FT		/product= unknown non-membrane protein
FT		/note= "gene E (specifically claimed)"
FT	CDS	7344..8897
FT		/*tag= f
FT		/product= putative efflux pump
FT		/note= "gene F (specifically claimed)"
FT	CDS	9164..10012
FT		/*tag= g
FT		/product= transcription activator
FT		/note= "gene G (specifically claimed)"
FT	CDS	complement (10105..10621)
FT		/*tag= h
FT		/product= translationally coupled to gene I
FT		/note= "gene H (specifically claimed)"
FT	CDS	complement (10618..11628)
FT		/*tag= i
FT		/product= homologue of fabH
FT		/note= "gene I (specifically claimed)"
FT	CDS	11809..12066
FT		/*tag= j
FT		/product= acyl carrier protein
FT		/note= "gene J (specifically claimed)"
FT	CDS	complement (12154..13209)
FT		/*tag= k
FT		/product= acyltransferase
FT		/note= "gene K (specifically claimed)"
FT	CDS	13409..14686
FT		/*tag= l
FT		/product= PKS ketoacylsynthase subunit
FT		/note= "gene L (specifically claimed)"
FT	CDS	14767..16047
FT		/*tag= m
FT		/product= PKS chain length factor
FT		/note= "gene M (specifically claimed)"
FT	CDS	16120..16371
FT		/*tag= n
FT		/product= acyl carrier protein used by the PKS
FT		/note= "gene N (specifically claimed)"
FT	CDS	complement (16453..16935)
FT		/*tag= o
FT		/product= putative hemiketal dehydrase
FT		/note= "gene O (specifically claimed)"
FT	CDS	17088..17903
FT		/*tag= p
FT		/product= ketoreductase related to acitli
FT		/note= "gene P (specifically claimed)"
FT	CDS	17903..18898
FT		/*tag= q
FT		/product= cyclase/dehydrase related to act VII
FT		/note= "gene Q (specifically claimed)"
FT	CDS	18895..19839
FT		/*tag= r

FT	/product= cyclase/dehydratase related to activ
FT	/note- "gene R (specifically claimed)"
FT	complement (19990..20907)
CDS	/tag= s
FT	/product= oxidoreductase
FT	/note- "gene S (specifically claimed)"
FT	complement (20904..22094)
CDS	/tag= t
FT	/product= quinone-forming hydroxylase
FT	/note- "gene T (specifically claimed)"
FT	complement (22505..22179)
CDS	/tag= u
FT	/product= not specified
FT	/note- "gene U"
FT	
X	

Dd	17567	cgtcgctcctgcgagcccccctaactgcgcgttccaagcagcgctgcgtcgttcaccaaagg	17626
Oy	576	gtcggtagcagagctgaacctgaagcctcaatcagcagcgcgctgcgccgcgtacat	635
Dd	17627	cctcgacctcgaaactgcgcgcgagcgcatacacgtcaacgcctctgcgcccgctacgt	17686
Oy	636	ccggcgcg	642
Dd	17687	cgagacg	17693

Search completed: August 19, 2002, 14:09:33  
Job time: 15272 sec

PN EP806480-A2.  
XX  
PD 12-NOV-1997.  
XX  
PF 02-MAY-1997; 97EP-0107329.  
XX  
PR 07-MAY-1996; 96US-0016753.  
XX  
XX  
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.  
XX  
XX  
PI Reeves CD, Soliday CL;  
XX  
XX WPI; 1997-538619/50.  
DR P-PSDB; AAMW34199-W34219.  
XX  
XX  
PT Streptomycetes frenolicin gene cluster - useful for producing  
PT recombinant frenolicin antibiotics  
XX  
XX  
PS Claim 1; Page 40-60; 66pp; English.

CC This DNA sequence comprises the Streptomyces freudenreichii gene  
CC cluster containing specifically claimed coding sequences (genes  
CC A-V) that respectively encode 21 proteins (see AMW34199-219) involved  
CC in farnicolin synthetase. The genes can be divided into 5  
CC subclusters: (1) genes A, B, C, D and E encode an efflux pump; (2)  
CC genes H, I, J and K encode butyrate starter synthetases; (3) genes  
CC L, M and N encode polyketide synthetases (PKS); (4) genes O, P, Q and  
CC R encode a hemiketalase, a ketoreductase and cyclases/dehydratases;  
CC and (5) genes S and T encode a keto/enoyl reductase and a  
CC hydrolase. Also claimed are vectors, host cells (especially a  
CC Streptomyces sp., particularly Streptomyces roseofulvus), and the  
CC encoded proteins. Cells transformed using the above sequence can  
CC be cultured to produce farnicolins or farnicolin precursors. The  
CC precursors can be converted to farnicolins by chemical or other  
CC methods. The farnicolins can be oxidised to farnicolin B, an  
CC antibiotic used as an anticoccidial agent. The farnicolins can be  
CC used as animal feed additives.

Query Match	6.5%	Score 47.8	DB 18	Length 24379
Best Local Similarity	47.2%	Pred. No. 0.54		
Matches 145	Conservative 0	Mismatches 162	Indels 0	Gaps 0

[illegible]

---

.....

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compgen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 13:45:36 ; Search time 130.83 Seconds  
(without alignments)  
1374.331 Million cell updates/sec

Title: US-10-003-759-3\_COPY\_52\_783

Perfect score: 732  
Sequence: 1 tgcgactgctcttccga.....atttccgaagctgag 732

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.NA:\*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/3B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCBUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131.8	18.0	1116	US-09-104-308-2	Sequence 2, Appl1
2	131.8	18.0	1116	US-09-321-981-2	Sequence 2, Appl1
3	105.8	14.5	1716	US-09-321-981-4	Sequence 4, Appl1
4	49	6.7	756	US-08-642-255-50	Sequence 50, Appl1
5	48.8	6.7	985	US-09-056-556-182	Sequence 182, Appl1
6	48	6.6	4403765	US-09-103-840A-2	Sequence 1, Appl1
7	47.6	6.5	30001	US-08-125-468-1	Sequence 1, Appl1
8	47.6	6.5	30001	US-08-474-933-1	Sequence 1, Appl1
9	47.2	6.4	1908	US-08-173-508-1	Sequence 1, Appl1
10	47.2	6.4	1908	US-08-265-310-1	Sequence 1, Appl1
11	47.2	6.4	1908	US-08-951-742-1	Sequence 1, Appl1
12	46.8	6.4	15872	US-09-105-537-1	Sequence 1, Appl1
13	46.2	6.3	1288	US-08-440-856A-9	Sequence 9, Appl1
14	46	6.3	44377	US-08-804-227C-7	Sequence 7, Appl1
15	46	6.3	44377	US-08-804-198-1	Sequence 1, Appl1
16	45.2	6.2	1291	US-08-997-897-1	Sequence 1, Appl1
17	45.2	6.2	1291	US-09-156-836B-1	Sequence 1, Appl1
18	44.6	6.2	2852	US-09-056-556-203	Sequence 203, Appl1
19	44.6	6.1	1176	US-08-387-942C-17	Sequence 17, Appl1
20	44.6	6.1	11230	US-09-105-537-32	Sequence 32, Appl1
21	44.6	6.1	12588	US-08-387-942C-1	Sequence 1, Appl1
22	44.6	6.1	36778	US-09-105-537-5	Sequence 5, Appl1
23	44.6	6.1	38506	US-09-320-878-19	Sequence 19, Appl1
24	44.6	6.1	2712	US-09-025-691-4	Sequence 4, Appl1
25	43.6	6.0	1017	US-08-849-751-1	Sequence 1, Appl1
26	43.6	6.0	1017	US-09-478-816-1	Sequence 1, Appl1
27	43.4	5.9	1500	US-09-593-711A-10	Sequence 10, Appl1

28	43	5.9	1273	4	US-09-319-892-3	Sequence 3, Appl1
29	43	5.9	1333	4	US-09-372-422A-9	Sequence 9, Appl1
30	42.8	5.8	599	1	US-08-584-226-3	Sequence 3, Appl1
31	42.4	5.8	20235	3	US-07-642-734C-3	Sequence 3, Appl1
32	42.4	5.8	20235	3	US-08-439-009A-3	Sequence 3, Appl1
33	42.2	5.8	1158	4	US-09-372-422A-21	Sequence 21, Appl1
34	42	5.7	71989	4	US-09-443-501A-2	Sequence 2, Appl1
35	41.8	5.7	432	1	US-08-642-255-48	Sequence 48, Appl1
36	41.6	5.7	1155	2	US-08-913-264-2	Sequence 2, Appl1
37	41.4	5.7	669	3	US-08-387-942C-7	Sequence 7, Appl1
38	41.4	5.7	18994	1	US-08-459-586-4	Sequence 4, Appl1
39	41.4	5.7	18994	2	US-08-282-696-4	Sequence 4, Appl1
40	41	5.6	1182	4	US-09-385-028-19	Sequence 19, Appl1
41	41	5.6	8051	2	US-08-576-626A-2	Sequence 2, Appl1
42	41	5.6	11604	4	US-09-385-028-13	Sequence 13, Appl1
43	41	5.6	15079	4	US-09-385-028-1	Sequence 1, Appl1
44	41	5.6	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
45	41	5.6	4411529	4	US-09-103-840A-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-09-104-308-2  
Sequence 2, Application US/09104308  
Patent No. 6187577  
GENERAL INFORMATION:  
APPLICANT: Jones, Brian E.  
APPLICANT: Van Der Kleij, Wilhelmus A.H.  
APPLICANT: Van Solingen, Piet  
TITLE OF INVENTION: No. 6187577e1 Cellulase Producing Actinomycetes,  
TITLE OF INVENTION: Cellulase Produced Therefrom and Method of Producing Same  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genecor International, Inc.  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09104, 308  
FILING DATE: 24-JUN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/974,042  
FILING DATE: 19-NOV-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Stone, Christopher L.  
REGISTRATION NUMBER: 35,696  
REFERENCE/DOCKET NUMBER: GC539  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7555  
TELEFAX: 650-846-6504  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1116 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-104-308-2

Query Match 18.0%; Score 131.8; DB 4; Length 1116;  
Best Local Similarity 50.7%; Pred. No. 2e-19;  
Matches 347; Conservative 0; Mismatches 332; Indels 6; Gaps 1;

```

OY      47      ccygagcccgagacgcgctgtgcaagcctgggaagcgcgatgtgccccgga          106
Db      77      CGGGCTCCGCCCGGAGCAACCGAGATGTGGCAGCCGCTTACGGCACACCACCATCCAGG          136
OY      107     ggcgcctaccgggtgatcaacaacgtaaggcgccggagaacccagtgcatgtagtgc          166
Db      137     ACCGGTAGATTGGTGCGAGAACCAACCCGTGGGGCGACCGGCACCCACCATGCATCAATGTGA          196
OY      167     gacttgaaacggygcaccttcagatcacacgcggccgatacgaacacgynaacaaogtgy          226
Db      197     CCGGCACCGCTTTCCGAGATCACCCAGCGCCGACGGCTTCGGTCCGACCAACAGGGCCCCGA          256
OY      227     ccgcattaccggscactcttcctcggtgcgaactggggcgcccttcacagagcaatlccggat          286
Db      257     AGTCCTATCCTCGGHTGCTACGAGCGGCTGCACACTAAGCAAACCTCGGGCCCCCGACGACGC          316
OY      287     tgcgcgcgagcgctgtcagagcgctgtccgaagtgcgcaecagcgctggaagcgtacgcgata          346
Db      317     TTCCCATATGGGATCAAGCTTCATCGGAGGGCGCCCGCCAGCAATGTCTCTTACCGGTACACCG          376
OY      347     cgaagcgccgcctygaaatgcgcgcctacagacatctgtgtcaagtcccgcgtcaagatltccgga          406
Db      377     GGAAGGGCGCTTACAACGGCGCGGTACGACATCTGG-----CTGGAAOCGCAACACCCCGCA          430
OY      407     acgycctacagcggcgagcgctgaatgaatcttgtcgtgaactvgaaocgsgcgctgaatgc          466
Db      431     CCAACGGGGGTGAACCGGACCGAGATCATGTGTTCAACCCGGGTGCGGCCGCTCAGC          490
OY      467     cgsgcgagcagcgcgctgtgcacccgtgaaactggcgggcgccacttggaaacttgttatg          526
Db      491     CCATCGCTTGGCGGCTGCGGACGCGCCACAGTGGCGGGCCCGCCACCTGGAGGTGTGGACCG          550
OY      527     ccgaactgggaactygaattacatcgtccacgcgcacagcgcaccaacgctcgtgtagcg          586
Db      551     GCAGCAACCGTTTCGAACGACGATGTCTCTTCTGGCGCCCTCCGGATCAGCAGCTGGA          610
OY      587     agctgagccttgaaagcctcatcogaagaagcggtgcgccggtctaactccgvcggagtc          646
Db      611     GCTTCGACGCTCAAGGACTTGTGACGACGCGCTCAAGCCAACGGCGCTGGCCACCCCGGACT          670
OY      647     ggtatcttgatgtcgtgtggaagcgggcttcgaactcttggaagggagggcgggcgctgcgaa          706
Db      671     GSTACTCTACACGACATCCAGCGGGGCTTCGAGCCGTGGAGCGGGGACCGGCTGTGGCGC          730
OY      707     gcgcgcgattttccgtaacggtgca          731
Db      731     TGAAGCTGTTCTCTCCGGCGTGA          755

RESULT      2
US-09-321-981-2
; Sequence 2, Application US/09321981
; Patent No. 6287839
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: NO. 6287839el Cellulase Producing Actinomyces,
; FILE REFERENCE: GS540-2
; CURRENT APPLICATION NUMBER: US/09/321,981
; PRIORITY FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: PCJ/US99/11971
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 09/104,308
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 08/974,042
; PRIOR FILING DATE: 1997-11-19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
```

```

; OTHER INFORMATION: Nearest "neighbor" - Streptomyces
; OTHER INFORMATION: thermovibrio
US-09-321-981-2

```

Query Match	18.0%;	Score 131.8;	DB 4;	Length 116;
Best Local Similarity	50.7%;	Pred. No. 2e-19;		
Matches 347;	Conservative	0;	Mismatches 332;	Indels 6; Gaps 1

[illegible]



```
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESS: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Mark David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 985 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-182

Query Match
Best Local Similarity 44.9%; Score 48.8; DB 4; Length 985;
Matches 267; Conservative 0; Mismatches 322; Indels 5; Gaps 2;

QY 13 ttcccgatgtagcaacaagaaagacccgagcccgagccgagccgagctgagctg 72
DB 175 ttccgagtgtagcccccgcacacagtgccgcccgcgcccgcgcccgcgcccgcg 234
QY 73 tgcgagcgtcgtagcgcgcgagctgtagccgagcgccgagccgagctgtagc 132
DB 235 accggtacccgagcgcgcgagctgtagccgagccgagccgagctgtagc 284
QY 133 tggagcgtagagacgcccagctgtagcagctgtagcagctgtagcagctg 132
DB 295 gccggcgccgcccgcgtagcagcgcgcgagtgagccgagccgagctgtagc 354
QY 193 acaagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 252
DB 355 ggcgtttgacgcccgcacagccgacagccgagccgagccgagccgagccgagc 414
QY 253 tgcagctgtagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 308
DB 415 atcaacggggcgccgagccgagccgagccgagccgagccgagccgagccgagc 474
QY 309 gtcgagcgtgtagcagcagctgtagcagcagcagcagcagcagcagcagcagc 368
DB 475 gccgcaggttccgagcaaacgagccgagccgagccgagccgagccgagccgagc 533
QY 369 ctacgagcagctgtagcagcagcagcagcagcagcagcagcagcagcagcagc 428
DB 534 cggcaacggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 593
QY 429 gtcgagcagctgtagcagcagcagcagcagcagcagcagcagcagcagcagc 488
DB 594 gggcaatggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 653
```

TREATM

```
QY 489 cgtgagctgcccgggagccacccctgggaagctgtagcagcagctgtagcagcagc 548
DB 654 cggcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 713
QY 549 cgcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 602
DB 714 caccgtatcgccgtaccgagccgagccgagccgagccgagccgagccgagccg 767

RESULT 6
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: ERASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match
Best Local Similarity 45.6%; Score 48; DB 4; Length 4403765;
Matches 208; Conservative 0; Mismatches 245; Indels 3; Gaps 1;

QY 100 gccgggggggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 159
DB 3963782 gttcgcaacggagacacccgacacacacacacacacacacacacacacacac 3963723
QY 160 gaagctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 219
DB 3963722 gttcgcaacggagacacccgacacacacacacacacacacacacacacacac 3963663
QY 220 aagctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 279
DB 3963662 atcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3963603
QY 280 tcgagctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 336
DB 3963602 atcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3963543
QY 337 acgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 396
DB 3963542 gttcgcaacggagacacccgacacacacacacacacacacacacacacacac 3963483
QY 397 aattccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 456
DB 3963482 aacaaacgttcgagcagcagcagcagcagcagcagcagcagcagcagc 3963423
QY 457 ggcgtgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 516
DB 3963422 ggcgaatgtagcagcagcagcagcagcagcagcagcagcagcagcagc 3963363
QY 517 gtcgtgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 552
DB 3963362 accggtttcgagcagcagcagcagcagcagcagcagcagcagcagcagc 3963327
```



```

RESULT 7
US-08-125-468-1/c
: Sequence 1, Application US/08125468
: Patent No. 5589385
: GENERAL INFORMATION:
:   APPLICANT: Ryan, Michael J.
:   APPLICANT: Lotvin, Jason A.
:   APPLICANT: Strathy, Nancy
:   APPLICANT: Fanti, Susan E.
:   TITLE OF INVENTION: Cloning of the biosynthetic pathway for
:   TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
:   NUMBER OF SEQUENCES: 1
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: American Cyanamid Company
:     STREET: One Cyanamid Plaza
:     CITY: Wayne
:     STATE: New Jersey
:     COUNTRY: USA
:     ZIP: 07470
:   COMPUTER READABLE FORM:
:     MEDIUM TYPE: Floppy disk
:     COMPUTER: IBM PC compatible
:     OPERATING SYSTEM: PC-DOS/MS-DOS
:     SOFTWARE: Patentin Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/125,468
:     FILING DATE: 22-SEP-1993
:     CLASSIFICATION: 435
:     ATTORNEY/AGENT INFORMATION:
:       NAME: Tsevdos, Estelle J
:       REGISTRATION NUMBER: 31,145
:       REFERENCE/DOCKET NUMBER: 31,255-02
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: (201)831-3241
:       TELEFAX: (201)831-3305
:     INFORMATION FOR SEO ID NO: 1:
:       SEQUENCE CHARACTERISTICS:
:         LENGTH: 30001 base pairs
:         TYPE: nucleic acid
:         STRANDEDNESS: single
:         TOPOLOGY: linear
:       MOLECULE TYPE: DNA (genomic)
:   US-08-125-468-1

Query Match          6.5%; Score 47.6; DB 1: Length 30001;
Best Local Similarity 45.3%; Pred No 0 076;
Matches 253; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

QY 144 gaccgcccagtcgattgagtcgagctgtaaaacgggcaacttcacgatcacagggccga 203
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 25022 GACCGCGCAAGCGCTTCGCCCTCGCGGCGAGGGCGCGCGGATGTTGTCTGAGAGCGGGA 24963

QY 204 tcacgacaacgggcaacaacagctggccgctatccggccatctacttcggtgtgccactggg 263
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24962 GCACCGCATCCCGCGCGGCGGCACACATCTACCGCGCGAGGTACCGCGGCGCTG 24903

QY 264 gcgcgcacagcaaatctgggagttggccggcgcgcgctgagggagctgtcgaagtgccgac 323
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24902 CACTCTCTTACACATGACCGGGCTGGGGTCTGCACGGGGAACGAATGACCGAGCTATCCG 24843

QY 324 gagctgagcagctcaacgcgcgatcatcagcagcgcgctgtgaaatgcgcgtacagacatctggt 383
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24842 GGTCT---GGGCTCGACCGGACGGCGGATGAACCCCGAGGACGTCGTACGTCACACCGCA 24786

QY 384 cagtcgcgttcacgaattccggaacggctacagcgcgcgcgccgagctgatgtatcttgct 443
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24785 CGGCTCGGCGACCA---GCAAGACGACCGGACGACGACGACCGCGCGCTTCAAGAGAGCCT 24729

QY 444 gaactcggaacggcgcggtgatgctcgcgcgcgcgccgctgtggccacccgtggaaactggccgg 503
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24728 GGGCGACCAACGCTTACCGGGTCCCGGTCACGCTTCGATCAAGTTCATGATCGGACACTCGCT 24669

```

```

OY      504      ggcacccgagggactgttgatcgcagctcggaagtcaattacatgcaccgcgcgcac   563
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      24668    GGGCCCGGTCTGCCGCTTGAGATTGCCGCCACAGCGGCTGGCATGCAGCACATCGGTGT   24609
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |

OY      564      gagcccaccacgcgtgtgagcgcagcttgcacctgaaggacctcatcgacaagcgcgttcgc   623
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      24608    GCCGCCACCAGGCCAAGCTTCACAGTGGCCCCGATCCCAGTAGCAGACTCACTGAATGCGCGCT   24549
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |

OY      624      ccggcgctcacatccgcgcggaatgtagtatgatgcgcgttgtgagaagcgcgttcgatactcg   683
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      24548    GGTTGGCGCGGAACAAGCGGGTGCAGCAGGTGCTCAGACGTGGCGAAGCGGCTTGGCGGCGCTT   24489
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      24488    CCAAGACCGCATGTGCT   24471


RESULT      8
US-08-474-933-1/c
Sequence 1, Application US/08474933
Patent No. 5866410
GENERAL INFORMATION:
APPLICANT: Ryan, Michael J.
APPLICANT: Lotvin, Jason A.
APPLICANT: Strachy, Nancy
APPLICANT: Fantini, Susan E.
TITLE OF INVENTION: Cloning of the biosynthetic pathway for
TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmids
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: USA
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC Compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,933
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,468
FILING DATE: 22-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Tsevdos, Estelle J
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,255-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3241
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match        6.5%; Score 47.6; DB 2; Length 30001;
Best Local Similarity 45.3%; Pred. No. 0.076;
Matches 253; Conservative 0; Mismatches 299; Indels 6; Gaps 2;

OY      144      gaccgccacgtgcatgtgagtcgagctgaaacaggccaacttcacgatcacacgggcgca   203
          |||| | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 25022 GACCCGACAGCGCTGCGCTCGGCGAGGCGCGCGATGTTGCTGGAGAGACCGGA 24963  
QY 204 tcacgaacaagcaagaacgtggccgacctatccggccatctacttggggcgcaactggg 263  
Db 24962 GCACGGGATCCGGCGCGCGCGCATCTACGCCGAGGTACCGGCTACCGCGCGCTG 24903  
QY 264 cgcctcgaagaacatcgggattcgccgagcggtgcaagagctgtcgcacg 323  
Db 24902 CACTCTACAGATGACCGGCTGCGGCTGACGAGGACAGATGCGCAGGCTATCCG 24843  
QY 324 gaactggaagctcaagcgctcaagaagcgctggaatgcgcgctcaacatctggt 383  
Db 24842 GGTG---GCGCTGACCGCGCGCGATGAGACCCCGAGAGCTGCGCTCAACCGCA 24786  
QY 384 caatccgctcaagatcgcgaagcgctacagcgagcgagcgagctgagttctgct 443  
Db 24785 CGGCTCGCGCACCA---GAGAGACGACCGCGACGAGACCGCGCGCTTCAAGAGACCT 24729  
QY 444 gaactggaagcgagcggtgagtcgagcgagcgagcgagcgagcgagcgagcgag 503  
Db 24728 GCGCGACACGCGCTACCGCGGCTGCGGCTGACGCTGATCAATGATGATCGACACTGCT 24669  
QY 504 ggcgaactggaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 563  
Db 24668 GGGCGCGCTGCGCGCTGAGTTGCGCGCGCGCGCGCTGCGCGCTGCGCGCTGCGGT 24609  
QY 564 gacgcgaacacgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 623  
Db 24608 GCG 24549  
QY 624 cgcgagctacatcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 683  
Db 24548 GGTGGCGCGCGAGGAGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 24489  
QY 684 ggaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 701  
Db 24488 CCAGAGCGCGATGCTGCT 24471

RESULT 9  
US-08-173-508-1  
; Sequence 1, Application US/08173508  
; Patent No. 5616485  
; GENERAL INFORMATION:  
; APPLICANT: Bartfeld, Daniel  
; APPLICANT: Butler, Michael J.  
; APPLICANT: Hadary, Dany  
; APPLICANT: Jenish, David  
; APPLICANT: Krieger, Timothy  
; TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED  
; TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/173,508  
; FILING DATE: 23-DEC-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 18740/125/CACO  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 202 672 5300  
TELEFAX: 202 672 5399  
TELEX: 904136  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1908 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 146..1759  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: 146..148  
OTHER INFORMATION: /note="Met at position -39  
represents fmet"  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 146..262  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 263..1756  
US-08-173-508-1

Query Match 6.4%; Score 47.2; DB 1; Length 1908;  
Best local Similarity 46.7%; Pred. No. 0.076;  
Matches 185; Conservative 0; Mismatches 208; Indels 3; Gaps 1;

QY 320 gcaagagctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 379  
Db 159 GCATACGGCGGAGGCGGAGCGCGCTTCGCGACGCGCGAGACACTGTCACCGCGCGCTGA 218  
QY 380 ggtcagtcgccgacgaatccggcaagcgagcgagcgagcgagcgagcgagcgagcgagcgag 439  
Db 219 TCGCCGCGCGCGCTGCG 278  
QY 440 ggtcagtcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 499  
Db 279 GCGCGAGCGTGGAGCGGAGCG 338  
QY 500 cgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 556  
Db 339 CGGCGATCGACCTGGAGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 398  
QY 557 ggcgaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 616  
Db 399 GCTACGTACAGGTGCGCGAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 458  
QY 617 cgtcgcgcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 676  
Db 459 TCGACCGCATCGCGACACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 518  
QY 677 aactcggagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 712  
Db 519 CCGCGGCTCCCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 554

RESULT 10  
US-08-265-310-1  
; Sequence 1, Application US/08265310  
; Patent No. 5856166  
; GENERAL INFORMATION:  
; APPLICANT: Bartfeld, Daniel  
; APPLICANT: Butler, Michael J.  
; APPLICANT: Hadary, Dany  
; APPLICANT: Jenish, David  
; APPLICANT: Krieger, Timothy  
; APPLICANT: Malek, Lawrence T.  
; APPLICANT: Soostmeyer, Gisela  
; APPLICANT: Walczyk, Eva  
; APPLICANT: Krzyznan, Phyllis



```

FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 263..1756
US-08-951-742-1

```

```

Query Match      6.4%; Score 47.2; DB 3; Length 1908;
Best Local Similarity 46.7%; Pred. No. 0.076;
Matches 185; Conservative 0; Mismatches 208; Indels 3; Gaps 1;

```

```

QY 320 gaacagagctggaagctcagcagccatcagcagcgccgctggaatgcccgaatcagatct 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 159 gcatacgggggaggggagcggccttggcagcggccgagacactggtcaccgccacctga 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 380 ggtcaagtcgcgtcagcaatcccgcaacggtctacagcgcgcgcgagcgtgatgatct 439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 219 tgcggcgccgctctctgcgaccccgccgagcgccgcccggcgccgagcgccagcgccagc 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 440 ggtcgaactcggaagcgcgctgcatgcccggcgcgagcccgctgtgcccacgttgaaactg 499
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 279 ggcggagctggagcggcgagcgccgctgcccctcccccgcggcgccgcccggcgccg 338
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 500 ccggggccacccggaagctgtatgcccgaactgggagctg---aatatcagcctacc 556
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 339 cggcgcatcactggagagcactgcgcagccgactggaacctgcccaagcccatcagatgcg 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 557 ggcgcacagcgcccaacacgctcgltgagcgagctggaactggaagcctcatalcgaagcg 616
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 399 gctacgtcacgggtgcgactgactacgccaagccgtaagcagcagatcagcgtcgccg 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 617 cggtcgcccgggtctacatccggcggaagtgtatctgcatcggtgaggaagcggtctcg 676
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 459 tgcagccgcatcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 677 aactctggaagcgcgagcggctctgcaagcgcg 712
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 519 cggcgcgctccggcgccgctccggcctctgcttcccg 554

```

```

RESULT 12
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

```

Query Match      6.4%; Score 46.8; DB 4; Length 15872;
Best Local Similarity 48.0%; Pred. No. 0.11;
Matches 172; Conservative 0; Mismatches 177; Indels 9; Gaps 1;

```

```

QY 289 ccggcgcgctgctcaggaagctgtcgcagcgcgcaagagctgtaagctcagccagatcag 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 764 cggcgcgagcggtactcgcgcgaaggtgcaagcgcgctgtaacctgcaagagctccacc 7705
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 349 aaggcgcgctgtaatgcgcctcagacatctgttcagtcgcgttcagaaatccggaac 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7706 aaggacgtcgcgctcgaagcctctgctctctctcctcgtcttcggaatcgatcgtagcacc 7765

```

```

QY 409 ggtcacagcgcgcgccagcagctgatgatctgtgctaactggaacggcggtgatgccc 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7766 gccgcagcgccaaactacgcgcgccaaacagcgctcagcgcctcgcgcgccacgcgc 7825
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 469 ggcggcgccggtggtccacccgtggaactggtcgcgggcgccacccgggaagtctgtatgcc 528
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7826 gccgcacgggctgtgcccgaacgctgtcgtcggtggtcgtcggtggaagcagcagcg 7885
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 529 gactggactggaatatactgcctacc-----ggcgacagcgcccaacacagctcg 579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7886 atggggagcagctcgcgccgcgcagcctcgcccgctgtgagccggcggaatcaccccg 7945
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 580 gtgagcgagctggaactgaagccttcacatcgaagaagcggtgcccggctacatcc 637
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7946 ctacaccgctcagagcgctcgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 8003

```

```

RESULT 13
US-08-440-856A-9
; Sequence 9, Application US/08440856A
; Patent No. 5750873
; GENERAL INFORMATION:
; APPLICANT: DELLAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVE. N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,856A
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLMAN, ROBERT A.
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 05463-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1517
; TELEFAX: (202) 887-0763
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1288 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-440-856A-9

```

```

Query Match      6.3%; Score 46.2; DB 1; Length 1288;
Best Local Similarity 38.5%; Pred. No. 0.12;
Matches 249; Conservative 48; Mismatches 346; Indels 4; Gaps 2;

```

```

QY 49 gagccgagccgacgctgagctgtgagcagctgtggaacgcgagtggtgcccgggg 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 339 gtcattgccgacatccacacgcsgcgsgagcgcctgcccgtggscggcgs 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 109 cgtacaggtgtatcaacaacatltggtgcgcggaagccgcacgaatgcatltgagtcgga 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 399 gtcrgsttgcgtggsgcagcagctgctgagagagagagagagagagagagagagagag 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 169 ctggaacggcgcaacttcagatcacagcgccgatcacagcgaacggaacaaagctgtgcc 228

```

DB 459 GCGCTGTSGCGSRGTGRCGCGSGS---CTSGACGTSTWGTGCAACACCCCGGGGTGCTG 515  
QY 229 gctatccggccatctactctggtgctgacactggtggtcctgacgaagcaatctcgatg 288  
DB 516 GCGCCGACAGACGGCGCGCGCGACAGACATCCTGTCTTGACGCGSSSAGTTGACGCC 575  
QY 289 ccggcgcgctgcaagaactctgctgcaagctgcaagactgcaagctcagccgcatcag 348  
DB 576 GTGCTTCGGGTCAAGCGCTGGGCGCGCTCGGSAATGAGACCGCGCGCGCATG 635  
QY 349 acggcgccgtgtaatgagccctcctacatctggtcagcccgctcagatccggaac 408  
DB 636 ACCCAGCGCGCGCGCG-CCGCGGAGACATCTCTCCGTGCGCACGCTCCSSGCTGCTG 694  
QY 409 ggtacacggcgcgcgccgacgtgatgactgctgactggaacggcggtgagcg 468  
DB 695 GCGCGCTGCGCGCGCGCGACGCTTACACCGGCTCCAGACGCGCATCTSSGGCTCACCAG 754  
QY 469 ggcggcagcgcgctggtgcaacgtggaactggtccggggccactggaagctggtatgc 528  
DB 755 AACGCGCGCTGCGAGCTSGCGSCGCGSRTCCGSGTCAACTGCTCTCSCCTTCGCG 814  
QY 529 gactggaactggaatatactgctcctacccggcgacgacgcccacgctcggtgagcg 588  
DB 815 GTGCGCACGCSATGCTCATCAACGCTGCGCGCACGCGCACGCMKCCACCGCGCGAC 874  
QY 589 ccggacgtgaagcctctacacgacgacggtgctgcccggcggtacacggcgagtg 648  
DB 875 GACGACCGAG 934  
QY 649 tatctgcatgctggtggaagcggtcgaactctggaagcggtggtg 695  
DB 935 AAGATGAG 981

RESULT 14  
US-08-804-227C-7/c  
Sequence 7, Application US/08804227C  
Patent No. 5876991

GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kunstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sulton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4437 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-227C-7

Query Match 6.3%; Score 46; DB 2; Length 44377;  
Best Local Similarity 47.6%; Pred. No. 0.17;  
Matches 136; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 19 gatggcgacacgaagaaagaaacccggaagcctgagccgagcgacgtcgagctgtcgga 78  
DB 18968 GACGCGGCGCTGCGCTTGGGCGGAGACACCGTGTGATCCGCTCGGCGCTGGCGGA 18909  
QY 79 cgttggagagcgcgatgtgctggtggggcgctacccggtgtacaaacgatatggcg 138  
DB 18908 GACGACCGCTGCTGAGCACCCTGGCGGTGTCACACGACGTCAGCGCGGTTCGT 18849  
QY 139 gcgagacgcccagtcgcatgagtgagctggaacggcaactcaacgatacaga 198  
DB 18848 GGGAGATGCTTCAGCAGCAGGCGCGGCGGCGGTCCGTACGTCAAGGCGCGTAC 18789  
QY 199 gccgatcagcaacgcaacgaacgctggtccgctatccggtcacttctggtggtccac 258  
DB 18788 GCGGACCTCGGACCGGACCGGCTGCAACTCGGCGGACAGTCAAGTCAAGCGCGGTGCTC 18729  
QY 259 tggggcgctgcaagcaatctggtggtggtggtggtggtggtggtggtggtggtggt 304  
DB 18728 GGGACCGCGCTGCGGACGACGAGAGATGCGCACCTGTGTGCGG 18683

RESULT 15  
US-08-804-198-1/c  
Sequence 1, Application US/08804198  
Patent No. 5945320

GENERAL INFORMATION:  
APPLICANT: Burgett, Stanley G.  
APPLICANT: Kunstoss, Stuart A.  
APPLICANT: Rao, Nagaraja R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Rostock, Paul R., Jr.  
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PAUL R. CANTRELL 1138  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 10:59:53 : Search time 3821.57 Seconds  
(without alignments)  
2585.265 Million cell updates/sec

Title: US-10-003-759-3\_COPY\_52\_783

Perfect score: 732  
Sequence: 1 tgcagctgcgtcttcccca.....atttcctgaacggtgcag 732

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inu:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrti:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	60.8	8.3	925	12	CNS0091P
2	56.2	7.7	925	12	CNS0091P
3	55.6	7.6	932	12	CNS00720
4	54.4	7.4	864	10	BF253875
5	53.8	7.3	964	12	CNS003WG
6	50.6	6.9	700	12	AG060149
7	50.4	6.9	935	12	CNS006XK
8	50.4	6.9	1022	9	BE040131
9	50	6.8	894	12	AC060214
10	50	6.8	1009	12	CNS010EW
11	49.8	6.8	449	10	BE517742
12	49.6	6.8	682	9	AU095988
13	49.2	6.7	534	12	AZ934730
14	49.2	6.7	546	12	AZ935386
15	48.2	6.6	443	9	AU030637
16	48.2	6.6	443	9	AV914669
17	48	6.6	446	10	BM373545

18	48	6.6	559	10	BE405069	BE405069	WHE1215-G
19	48	6.6	588	10	BF255157	BF255157	HVSMF000
20	48	6.6	604	9	AV941745	AV941745	AV941745
21	48	6.6	606	10	BF255408	BF255408	HVSMF000
22	48	6.6	631	9	AL505254	AL505254	AL505254
23	48	6.6	635	9	AV932837	AV932837	AV932837
24	48	6.6	637	9	AV933847	AV933847	AV933847
25	48	6.6	666	10	BF628036	BF628036	HVSMF000
26	48	6.6	719	9	AV935472	AV935472	AV935472
27	48	6.6	830	10	BF254391	BF254391	HVSMF000
28	48	6.6	844	12	CNS0052P	AL056652	Drosophila
29	48	6.6	1101	12	CNS017SY	AL108460	Drosophila
30	47.8	6.5	307	9	AU070991	AU070991	AU070991
31	47.8	6.5	578	9	AU094712	AU094712	AU094712
32	47.6	6.5	623	9	AU032635	AU032635	AU032635
33	47.6	6.5	517	9	AV933848	AV933848	AV933848
34	47.6	6.5	935	12	CNS006XK	AL066051	Drosophila
35	47.4	6.5	407	10	BE475922	BE475922	946049606
36	47	6.4	657	10	BE17319	BE17319	1031019B1
37	47	6.4	677	9	AL505169	AL505169	AL505169
38	46.8	6.4	269	10	BE636900	BE636900	WHE1805-1
39	46.8	6.4	641	9	AU093820	AU093820	AU093820
40	46.8	6.4	669	9	AU096005	AU096005	AU096005
41	46.8	6.4	674	9	AU055758	AU055758	AU055758
42	46.6	6.4	691	9	AL504955	AL504955	AL504955
43	46.6	6.4	839	12	CNS004NB	AL054280	Drosophila
44	46.4	6.3	453	10	BE1779131	BE1779131	EBRC01_SO
45	46.4	6.3	557	10	BF618612	BF618612	HVSMF000

## ALIGNMENTS

RESULT 1  
LOCUS 925 bp DNA linear GSS 03-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACRI9D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL053013  
AL053013.1 GI:4934461  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 925)

## REFERENCE

Genoscope.  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

## COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Hammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source  
1..925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"







[illegible]

```

Db      834  ANMSSGSGSGSSGGSVSRTRTKRRRSSGGBSGGKSGGGRGAAGCGGGGGRGG 775
Oy      466  ccggagcggcagccgcgctgcccacgctggaactgcccggggccacactgggaagctgatat 525
Db      774  SRSSTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 715
Oy      526  gccgcactgggaactgaattacatgcctccacccggcgacgacagcagccacacagctgcgtgac 585
Db      714  GSTTSSKVGSKTCTTSGKVNNTSSSSYCRSTNSGABRTBTBSAANAADNNS -SAMSMMC 656
Oy      586  gagctggaactgaagcctcattcgcagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 645
Db      655  SASSVVGVCSCSCMSCASSSCSSSCCMVCCGCGKCFKCSSSVSRSGSRTGCGVCGCGSSG 596
Oy      646  tggatctgcgcatcggtggagagcgcttcgaactctgggaaggagggcggtgcgtctgcga 705
Db      595  SGBTTSSGCGVAAAKACACACGACGKAGGSSVSGCCCTGTWSSRADCAAGMACRRVSSGTIVA 536
Oy      706  agcgcc 711
Db      535  VABTSH 530

RESULT 6
LOCUS   AG060149
DEFINITION Pan troglodytes DNA, clone: PTB-047122.R, genomic survey sequence.
ACCESSION AG060149
KEYWORDS  AG060149.1 GI:16611379
SOURCE   GSS: (genome survey sequence).
          Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
          BAC library clone:PTB-047122.R.
          Pan troglodytes
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

ORGANISM
REFERENCE
AUTHORS  Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          ToTokI,Y., Matanabe,H. and Sakaki,Y.
TITLE    BAC end sequences of library PTB
REFERENCE Unpublished
AUTHORS  Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
          ToTokI,Y., Matanabe,H. and Sakaki,Y.
TITLE    Direct Submission
JOURNAL  Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suenhiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail:chimbess@gsc.riken.go.jp, URL:http://bgp.gsc.riken.go.jp/,
          Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT  Clones are derived from the chimpanzee BAC library PTB This BAC end
          clone was generated during the R&D process and may have higher chance of
          clone tracking errors.
          PRIMERS
          Sequencing: M13Rev
          LIBRARY
          Vector : pKS145
          R.Site 1 : SacI
          R.Site 2 : SacI.
          Location/Qualifiers
            1..700
              /organism="Pan troglodytes"
              /db_xref="taxon:9598"
              /clone="PTB-047122.R"
              /sex="male"
              /cell_type="lymphoblast"
              /clone_lib="PTB Chimpanzee Male BAC Library"
              30 a 291 c 334 g 15 t 30 others

BASE COUNT
ORIGIN
Query Match 6.9%; Score 50.6; DB 12; Length 700;
Best Local Similarity 42.7%; Pred. NO. 11;
Matches 243; Conservative 0; Mismatches 323; Indels 3; Gaps 1;

```







LOCUS AU095988 682 bp mRNA linear EST 30-JUN-2000  
 DEFINITION AU095988 Rice green shoot *Oryza sativa* cDNA clone S11827, mRNA  
 sequence.  
 ACCESSION AU095988  
 VERSION AU095988.1 GI:8558670  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM *Oryza sativa*.  
*Oryza sativa*.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; *Oryza*.  
 REFERENCE 1 (bases 1 to 682)  
 AUTHORS Sasaki, T. and Yamamoto, K.  
 TITLE Rice cDNA from green shoot (2000)  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abr.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/  
 PROJECT "RGP".  
 S11827\_102.

FEATURES  
 source  
 1..682  
 /organism="Oryza sativa"  
 /strain="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="S11827"  
 /clone\_lib="Rice green shoot"  
 /note="Green shoot (8 days old)"  
 /note="Green shoot (8 days old)"  
 153 a 184 c 212 g 132 t 1 others

BASE COUNT 153 a 184 c 212 g 132 t 1 others  
 ORIGIN

Query Match 6.88; Score 49.6; DB 9; Length 682;  
 Best Local Similarity 47.1%; Pred. No. 16;  
 Matches 185; Conservative 0; Mismatches 205; Indels 3; Gaps 1;

248 tggggggccacttgaggcgcttgacgagcaattcgatgcgcggcgcgctgcaagagc 307  
 Db 453 TGGCCTCTCTCGAGAGCTGGTCTTCGACAGACGCCGAGTCTTGTTCACAGTCCAGTACA 394  
 QY 308 tgcgcagctgacgagcgtcagcgtcagcgcgacacacgagcagcgcgctggaatcgcg 367  
 Db 393 CGCGGAGCTTGTCAGAGCTCGGGTGGCCGACACACGCTGCGCAGCGCTTCCGGAGTCCG 334  
 QY 368 cctcagacatctggttcagtcctcagcaattccgagcaagcgctacagcgcgcgcg 427  
 Db 333 CCTCCACACAGAACCCCTTCTGTCACGAAACGCTGGAACGGCGGGAGACGACGCCGGAACA 274  
 QY 428 agctgatgacatcgctggaacttgaaacgcgctgatatcgcgcgcgagcgcgctgagcca 487  
 Db 273 GCGGAGATGTCTCGCGAACAACCCCTGTCGCGATGCAACCCGGGTTACAGCAGCCGGA 214  
 QY 488 ccgttggaacttgccgggggccaacctgggaagctctgtaatacgcaactgggaactggaataca 547  
 Db 213 ACGTGATCCCGGTCT---CCTCGTGAATCTCGGTGAGACCTCTGATCGTACAGCATGT 157  
 QY 548 tcgctcaccggcgagcaagcaccacacgctcggtgagcgagctggaactgaagccttca 607  
 Db 156 TACAGATCTTGTGCTGTCTTGTACGCCCTTGCGCGCCGTGCAAGCTCTCGCGCCGTGATCA 97  
 QY 608 tcgacgagcggtcgccgcgagctacatccgcg 640  
 Db 96 TCGCCGACCCGTTCTGCGCCCGGAGACGCCGCGC 64

RESULT 13  
 A2934730/c 534 bp DNA linear GSS 24-APR-2001  
 LOCUS  
 DEFINITION BT\_Ba0002f24r B. japonicum BAC library Bradyrhizobium japonicum

genomic, DNA sequence.  
 ACCESSION A2934730  
 VERSION A2934730.1 GI:13776790  
 KEYWORDS GSS.  
 SOURCE Bradyrhizobium japonicum.  
 ORGANISM Bradyrhizobium japonicum.  
 Bradyrhizobium japonicum  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Bradyrhizobium group; Bradyrhizobium.  
 REFERENCE 1 (bases 1 to 534)  
 AUTHORS Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Golcochea  
 J.L., Stacey, G., Sadowsky, M.J. and Wing, R.A.  
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum  
 genome  
 JOURNAL Genome Res. 11 (8), 1434-1440 (2001)  
 MEDLINE 21376150  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Class: BAC ends  
 High quality sequence stop: 497.  
 Location/Qualifiers  
 1..534  
 /organism="Bradyrhizobium japonicum"  
 /strain="USD110"  
 /db\_xref="taxon:375"  
 /clone\_lib="B. japonicum BAC library"  
 /lab\_host="E. coli"  
 /note="Vector: pindig9536; site\_1: HindIII"  
 99 a 170 c 185 g 80 t

BASE COUNT 99 a 170 c 185 g 80 t  
 ORIGIN

Query Match 6.7%; Score 49.2; DB 12; Length 534;  
 Best Local Similarity 51.4%; Pred. No. 19;  
 Matches 114; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

262 gggccctgcacggagcaattcgatgctgcgcggcgcggtgaggaagctgtccgacgtgcg 321  
 Db 498 GGCACCCGCGATCTGCTTCGATCCGATCGCGGTGCGCGGTCGACATCTACGTCATGCGC 439  
 QY 322 acgagctgacgtcaccgcgacacgagcagcgcgcttggaatgcgcgctacgacatctgg 381  
 Db 438 GCGGGGGGGAGACTGGCAGCGCATCTCTTCTCAAGAGACACACAAAGCCACTTAT 379  
 QY 382 ttaagtcgcgtcacgaattccgcaacgagctacagcgcgcgcgagctgatatctgg 441  
 Db 378 TCGACGCGGCTGTGTCGCGCGCGCGGCGACTACATGCGCTTCACAGGCAAGGCGCGGA 319  
 QY 442 ctgaactggaacgcgcgctgacgcggcgcgcgagcgagcgctg 483  
 Db 318 CAGTTTCATCGCGGCTGATGAAGCCGGACGCGAGCGCGGANG 277

RESULT 14  
 A2935386 546 bp DNA linear GSS 24-APR-2001  
 LOCUS  
 DEFINITION BT\_Ba0003b10r B. japonicum BAC library Bradyrhizobium japonicum  
 genomic, DNA sequence.  
 ACCESSION A2935386  
 VERSION A2935386  
 KEYWORDS GSS.  
 SOURCE Bradyrhizobium japonicum.  
 ORGANISM Bradyrhizobium japonicum.  
 Bradyrhizobium japonicum  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Bradyrhizobium group; Bradyrhizobium.  
 REFERENCE 1 (bases 1 to 546)  
 AUTHORS Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Golcochea  
 J.L., Stacey, G., Sadowsky, M.J. and Wing, R.A.  
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum



